#### **PCT**

### WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



### INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

WO 00/11157 (51) International Patent Classification 7: (11) International Publication Number: C12N 15/11, 15/00, C07K 16/00 **A1** 2 March 2000 (02.03.00) (43) International Publication Date: PCT/US99/19395 (21) International Application Number:

(30) Priority Data:

(22) International Filing Date:

60/097,927

25 August 1998 (25.08.98) US

25 August 1999 (25.08.99)

(71) Applicant: THE GENERAL HOSPITAL CORPORATION [US/US]; 55 Fruit Street, Boston, MA 02114 (US).

(72) Inventors: BROWN, Robert, H.; 16 Oakland Avenue, Needham, MA 02192 (US). LIU, Jing; 1629 Van Horn, Outremont, Quebec H2V IL1 (CA). AOKI, Masashi; Dept. of Neurology, Tohoku University School of Medicine, 1-1, Seiro-Machi, Aoba-ku, Sendai (JP). HO, Meng, F.; Apartment 24, 145 Englewood Avenue, Brighton, MA 02135 (US). MATSUDA-ASADA, Chie; 33 Pond Avenue, Brookline, MA 02445 (US).

(74) Agent: FRASER, Janis, K.; Fish & Richardson P.C., 225 Franklin Street, Boston, MA 02110-2804 (US).

(81) Designated States: CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT,

#### Published

With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: DYSFERLIN, A GENE MUTATED IN DISTAL MYOPATHY AND LIMB GIRDLE MUSCULAR DYSTROPHY

MM candidate region

# Centromere Telomere Previous candidate region

#### (57) Abstract

A novel gene and the protein encoded therein, i.e., dysferlin, are disclosed. This gene and its expression products are associated with muscular dystrophy, e.g., Miyoshi myopathy and limb girdle muscular dystrophy 2B.

#### FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	ТJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TТ	Trinidad and Tobago
BJ	Benin	1E	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Кепуа	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	zw	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

BNSDOCID: <WO\_\_\_\_\_0011157A1\_I\_>

## DYSFERLIN, A GENE MUTATED IN DISTAL MYOPATHY AND LIMB GIRDLE MUSCULAR DYSTROPHY

#### RELATED APPLICATION INFORMATION

This application claims priority from provisional application serial no. 60/097,927, filed August 25, 1998.

#### Statement as to Federally Sponsored Research

The work described herein was supported in part by 10 NIH grants 5P01AG12992, 5R01N834913A, and 5P01NS31248.

The Federal Government therefore may have certain rights in the invention.

#### Background of the Invention

The invention relates to genes involved in the 15 onset of muscular dystrophy.

Muscular dystrophies constitute a heterogeneous group of disorders. Most are characterized by weakness and atrophy of the proximal muscles, although in rare myopathies such as "Miyoshi myopathy" symptoms may first 20 arise in distal muscles. Of the various hereditary types of muscular dystrophy, several are caused by mutations or deletions in genes encoding individual components of the dystrophin-associated protein (DAP) complex. It is this DAP complex that links the cytoskeletal protein 25 dystrophin to the extracellular matrix protein, laminin-2.

Muscular dystrophies may be classified according to the gene mutations that are associated with specific clinical syndromes. For example, mutations in the gene encoding the cytoskeletal protein dystrophin result in either Duchenne's Muscular Dystrophy or Becker's Muscular Dystrophy, whereas mutations in the gene encoding the extracellular matrix protein merosin produce Congenital

5

- 2 -

Muscular Dystrophy. Muscular dystrophies with an autosomal recessive mode of inheritance include "Miyoshi myopathy" and the several limb-girdle muscular dystrophies (LGMD2). Of the limb-girdle muscular dystrophies, the deficiencies resulting in LGMD2C, D, E, and F result from mutations in genes encoding the membrane-associated sarcoglycan components of the DAP complex.

#### Summary of the Invention

A novel protein, designated dysferlin, is 10 identified and characterized. The dysferlin gene is normally expressed in skeletal muscle cells and is selectively mutated in several families with the hereditary muscular dystrophies, e.g., Miyoshi myopathy 15 (MM) and limb girdle muscular dystrophy-2B (LGMD2B). These characteristics of dysferlin render it a candidate disease gene for both MM and LGMD2B. An additional novel protein, brain-specific dysferlin, has also been identified. Defects in brain-specific dysferlin may 20 predispose to selected disorders of the central nervous system. Moreover, the expression of brain-specific dysferlin may be important as a marker for normal neural development (e.g., in vivo or in neural cells in culture). Manipulation of levels of expression of brain-25 specific dysferlin, and of the type of expressed brainspecific dysferlin is of use for analyzing the function of brain-specific dysferlin and related dysferlinassociated molecules.

The invention features an isolated DNA which

30 includes a nucleotide sequence hybridizing under

stringent hybridization conditions to a strand of SEQ ID

NO:3 or SEQ ID NO:117.

The invention also features an isolated DNA including a nucleotide sequence selected from SEQ ID NOs:4-12.

Also within the invention is an isolated DNA comprising a nucleotide sequence selected from the group consisting of SEQ ID NOs:22-30.

Also within the invention is a single stranded oligonucleotide of 14-50 nucleotides in length having a nucleotide sequence identical to a portion of a strand of 10 SEQ ID NO:3.

Also within the invention is a pair of PCR primers consisting of:

- (a) a first single stranded oligonucleotideconsisting of 14-50 contiguous nucleotides of the sense15 strand of SEQ ID NO:117; and
- (b) a second single stranded oligonucleotide consisting of 14-50 contiguous nucleotides of the antisense strand of SEQ ID NO:117, wherein the sequence of at least one of the oligonucleotides is identical to a 20 portion of a strand of SEQ ID NO:3, and the first oligonucleotide is not complementary to the second oligonucleotide.

Also within the invention is a pair of single stranded oligonucleotides selected from of SEQ ID NO: 130-231, SEQ ID NO:110, and SEQ ID NO:112.

Also within the invention is an isolated DNA including a nucleotide sequence that encodes a protein that shares at least 70% sequence identity with SEQ ID NO:2, or a complement of the nucleotide sequence.

Also within the invention is an isolated DNA including a nucleotide sequence which hybridizes under stringent hybridization conditions to a strand of a nucleic acid, the nucleic acid having a sequence selected from SEQ ID NOs:31-79 and 90-101.

- 4 -

Also within the invention is a single stranded oligonucleotide of 14-50 nucleotides in length having a nucleotide sequence which is identical to a portion of a strand of a nucleic acid selected from SEQ ID NOs:31-79 and 90-100.

Also within the invention is a pair of PCR primers consisting of:

- (a) a first single stranded oligonucleotide consisting of 14-50 contiguous nucleotides of the sensestrand of a nucleic acid selected from SEQ ID NOs:31-85; and
- (b) a second single stranded oligonucleotide consisting of 14-50 contiguous nucleotides of the antisense strand of a nucleic acid selected from SEQ ID NOs:31-85, wherein the sequence of at least one of the oligonucleotides includes a sequence identical to a portion of a strand of a nucleic acid selected from SEQ ID NOs: 31-79 and 90-100, and the first oligonucleotide is not complementary to the second oligonucleotide.
- Also within the invention is a pair of single stranded oligonucleotides selected from SEQ ID NOs 101-116, SEQ ID NOs 184-185, SEQ ID NOs 188-191, SEQ ID NOs 210-213, and SEQ ID NOs 216-217.

Also within the invention is a substantially pure protein that has an amino acid sequence sharing at least 70% sequence identity with SEQ ID NO:2.

Also within the invention is a substantially pure protein the sequence of which includes amino acid residues 1-500, 501-1000, 1001-1500, or 1501-2080 of SEQ 30 ID NO:2.

Also within the invention is a substantially pure protein including the amino acid sequence of SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, or SEQ ID NO:89.

In another aspect, the invention features a 35 transgenic non-human mammal having a transgene disrupting

PCT/US99/19395 WO 00/11157

- 5 -

or interfering with the expression of a dysferlin gene, the transgene being chromosomally integrated into the germ cells of the animal.

Another embodiment of the invention features a 5 method of decreasing the symptoms of muscular dystrophy in a mammal by introducing into a cell of the mammal (e.g., a muscle cell or a muscle precursor cell) an isolated DNA which hybridizes under stringent hybridization conditions to a strand of SEQ ID NO:3.

Another aspect of the invention provides a method for identifying a patient, a fetus, or a pre-embryo at risk for having a dysferlin-related disorder by (a) providing a sample of genomic DNA from the patient, fetus, or pre-embryo; and (b) determining whether the 15 sample contains a mutation in a dysferlin gene.

In another aspect, the invention provides a method for identifying a patient, a fetus, or a pre-embryo at risk for having a dysferlin-related disorder by (a) providing a sample including dysferlin mRNA from the 20 patient, fetus, or pre-embryo; and (b) determining whether the dysferlin mRNA contains a mutation.

Methods of identifying mutations in a dysferlin sequence are useful for predicting (e.g., predicting whether an individual is at risk for developing a 25 dysferlin-related disorder) or diagnosing disorders associated with dysferlin, e.g., MM and LGMD2B. Such methods can also be used to determine if an individual, fetus, or a pre-embryo is a carrier of a dysferlin mutation, for example in screening procedures. Methods 30 which distinguish between different dysferlin alleles (e.g., a mutant dysferlin allele and a normal dysferlin allele) can be used to determine carrier status.

The invention also features an isolated nucleic acid comprising a nucleotide sequence which hybridizes 35 under stringent hybridization conditions to nucleic acids

10

- 6 -

3284-3720 of SEQ ID NO:232, or the complement of the nucleotide sequence. An isolated nucleic acid including a nucleotide sequence identical to the sequence of nucleotides 3284-3720 of SEQ ID NO:232, or a complement of the nucleotide sequence is also a feature of the invention. The isolated nucleic acid can include the entire sequence of SEQ ID NO:232 or the complement of SEQ ID NO:232.

Another aspect of the invention features an isolated polypeptide that includes: a) at least 15 contiguous amino acids of the polypeptide comprising amino acids 1-24 of SEQ ID NO:233, b) a naturally occuring allelic variant of a polypeptide comprising amino acids 1-24 of SEQ ID NO:233, or c) an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes under stringent conditions to nucleotides 3284-3720 of SEQ ID NO:232. The polypeptide of this aspect can include the entire sequence of SEQ ID NO:233.

Also included in the invention is a vector comprising the nucleic acid of claim 44 and a cell that contains the vector. Another aspect of the invention features a method of making a polypeptide by culturing the cell which contains the vector.

The invention also features an antibody which specifically binds to a polypeptide of such as those described above. The antibody can bind to a polypeptide selected from amino acids 253-403 of SEQ ID NO:233, amino acids 624-865 of SEQ ID NO:233, and amino acids 1664-1786 of SEQ ID NO:233. Antibodies of the invention can be monclonal or polyclonal antibodies.

An "isolated DNA" is DNA which has a naturally occurring sequence corresponding to part or all of a given gene but is free of the two genes that normally flank the given gene in the genome of the organism in

- 7 -

which the given gene naturally occurs. The term
therefore includes a recombinant DNA incorporated into a
vector, into an autonomously replicating plasmid or
virus, or into the genomic DNA of a prokaryote or
eukaryote. It also includes a separate molecule such as
a cDNA, a genomic fragment, a fragment produced by
polymerase chain reaction (PCR), or a restriction
fragment, as well as a recombinant nucleotide sequence
that is part of a hybrid gene, i.e., a gene encoding a
fusion protein. The term excludes intact chromosomes and
large genomic segments containing multiple genes
contained in vectors or constructs such as cosmids, yeast
artificial chromosomes (YACs), and P1-derived artificial
chromosome (PAC) contigs.

A "noncoding sequence" is a sequence which corresponds to part or all of an intron of a gene, or to a sequence which is 5' or 3' to a coding sequence and so is not normally translated.

An expression control sequence is "operably
20 linked" to a coding sequence when it is within the same
nucleic acid and can control expression of the coding
sequence.

A "protein" or "polypeptide" is any chain of amino acids linked by peptide bonds, regardless of length or post-translational modification, e.g., glycosylation or phosphorylation.

As used herein, the term "percent sequence identity" means the percentage of identical subunits at corresponding positions in two sequences when the two sequences are aligned to maximize subunit matching, i.e., taking into account gaps and insertions. For purposes of the present invention, percent sequence identity between two polypeptides is to be determined using the Gap program and the default parameters as specified therein.

35 The Gap program is part of the Sequence Analysis Software

BNSDOCID: <WO\_\_\_\_0011157A1\_I\_>

- 8 -

Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705.

The algorithm of Myers and Miller, CABIOS (1989)

5 can also be used to determine whether two sequences are similar or identical. Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a

10 PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used.

As used herein, the term "stringent hybridization conditions" means the following DNA hybridization and wash conditions: hybridization at 60°C in the presence of 6 x SSC, 0.5% SDS, 5 x Denhardt's Reagent, and 100 µg/ml denatured salmon sperm DNA; followed by a first wash at room temperature for 20 minutes in 0.5 x SSC and 0.1% SDS and a second wash at 55°C for 30 minutes in 0.2 x SSC and 0.1% SDS.

A "substantially pure protein" is a protein 20 separated from components that naturally accompany it. The protein is considered to be substantially pure when it is at least 60%, by dry weight, free from the proteins and other naturally-occurring organic molecules with 25 which it is naturally associated. Preferably, the purity of the preparation is at least 75%, more preferably at least 90%, and most preferably at least 99%, by weight. A substantially pure dysferlin protein can be obtained, for example, by extraction from a natural source, by 30 expression of a recombinant nucleic acid encoding a dysferlin polypeptide, or by chemical synthesis. Purity can be measured by any appropriate method, e.g., column chromatography, polyacrylamide gel electrophoresis, or HPLC analysis. A chemically synthesized protein or a 35 recombinant protein produced in a cell type other than

PCT/US99/19395 WO 00/11157

- 9 -

the cell type in which it naturally occurs is, by definition, substantially free from components that naturally accompany it. Accordingly, substantially pure proteins include those having sequences derived from 5 eukaryotic organisms but which have been recombinantly produced in E. coli or other prokaryotes.

An antibody that "specifically binds" to an antigen is an antibody that recognizes and binds to the antigen, e.g., a dysferlin polypeptide, but which does 10 not substantially recognize and bind to other molecules in a sample (e.g., a biological sample) which naturally includes the antigen, e.g., a dysferlin polypeptide. antibody that "specifically binds" to dysferlin is sufficient to detect a dysferlin polypeptide in a 15 biological sample using one or more standard immunological techniques (for example, Western blotting or immunoprecipitation).

A "transgene" is any piece of DNA, other than an intact chromosome, which is inserted by artifice into a 20 cell, and becomes part of the genome of the organism which develops from that cell. Such a transgene may include a gene which is partly or entirely heterologous (i.e., foreign) to the host organism, or may represent a gene homologous to an endogenous gene of the organism.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Methods and materials similar or equivalent to those described herein can be 30 used in the practice or testing of the present invention. The present materials, methods, and examples are illustrative only and not intended to be limiting. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference 35 in their entirety. In case of conflict, the present

25

- 10 -

specification, including definitions, will control. All the sequences disclosed in the sequence listing are meant to be double-stranded except the sequences of oligonucleotides.

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

#### Brief Description of the Drawings

Fig. 1A is a physical map of the MM locus. Arrows indicate the five new polymorphic markers and filled, vertical rectangular boxes indicate the previously known polymorphic markers. The five ESTs that are expressed in skeletal muscle are highlighted in bold. Detailed information on the minimal tiling path of the PAC contig spanning the MM/LGMD2B region is provided in Liu et al., 1998, Genomics 49:23-29. The minimal candidate MM region is designated by the solid bracket (top) and compared to the previous candidate region (dashed bracket). TGFA and ADD2 are transforming growth factor alpha and  $\beta$ -adducin 20 2.

Fig. 1B is a representation of the dysferlin cDNA clones. The probes used in the three successive screens are shown in bold (130347, cDNA10, A27-F2R2). The two most 5' cDNA clones are also shown (B22, B33). The 6.9 25 kb cDNA for dysferlin (SEQ ID NO:1) is illustrated at the bottom with start and stop codons as shown.

Fig. 1C is a representation of the predicted dysferlin protein. The locations of four C2 domains (SEQ ID NOs: 86-89) are indicated by stippled boxes,

while the putative transmembrane region is hatched.

Vertical lines above the cDNA denote the positions of the mutations in Table 2; the associated labels indicate the phenotypes (MM - Miyoshi myopathy; LGMD - limb girdle

- 11 -

muscular dystrophy; DMAT - distal myopathy with anterior tibial onset).

Fig. 2 is the sequence of the predicted 2,080 amino acids of dysferlin (SEQ ID NO:2). The predicted 5 membrane spanning residues are in bold at the carboxy terminus (residues 2047-2063). Partial C2 domains are underlined. Bold, underlined sequences are putative nuclear targeting residues. Possible membrane retention sequences are enclosed within a box.

10 Fig. 3 is a comparison of the Kyle-Doolittle hydrophobicity plots of the dysferlin protein and fer-1. On the Y-axis, increasing positivity corresponds to increasing hydrophobicity. Both proteins have a single, highly hydrophobic stretch at the carboxy terminal end (arrow). Both share regions of relative hydrophilicity approximately at residue 1,000 (arrowhead).

Fig. 4 is a SSCP analysis of a representative pedigree with dysferlin mutations. Each member of the pedigree is illustrated above the corresponding SSCP analysis. For each affected individual (solid symbols) shifts are evident in alleles 1 and 2, corresponding respectively to exons 36 and 54. As indicated, the allele 1 and 2 variants are transmitted respectively from the mother and the father. The two affected daughters in this pedigree have the limb girdle muscular dystrophy (LGMD) phenotype while their affected brother has a pattern of weakness suggestive of Miyoshi myopathy (MM).

Fig. 5 is a representation of the genomic structure of dysferlin. The 55 exons of the dysferlin 30 gene and their corresponding SEQ ID NOs are indicated below the 6911 bp cDNA (solid line). The cDNA sequences corresponding to SEQ ID NO:1 and SEQ ID NO:3 are shown relative to the 6911 bp cDNA.

Figs. 6A-B are the cDNA sequence of brain-specific 35 dysferlin (SEQ ID NO:232) and the predicted amino acid

- 12 -

sequence (in single-letter code) of brain-specific dysferlin (SEQ ID NO:233).

#### <u>Detailed Description</u>

The Miyoshi myopathy (MM) locus maps to human 5 chromosome 2p12-14 between the genetic markers D2S292 and D2S286 (Bejaoui et al., 1995, Neurology 45:768-72). Further refined genetic mapping in MM families placed the MM locus between markers GGAA-P7430 and D2S2109 (Bejaoui et al., 1998, Neurogenetics 1:189-96). Independent 10 investigation has localized the limb-girdle muscular dystrophy (LGMD-2B) to the same genetic interval (Bashir et al., 1994, Hum. Molec. Genetics 3:455-57; Bashir et al., 1996, Genomics 33:46-52; Passos-Bueno et al., 1995, Genomics 27:192-95). Furthermore, two large, inbred .15 kindreds have been described whose members include both MM and LGMD2B patients (Weiler et al., 1996, Am. J. Hum. Genet. 59:872-78; Illarioshkin et al., 1997, Genomics 42:345-48). In these familial studies, the disease gene(s) for both MM and LGMD2B mapped to essentially the 20 same genetic interval. Moreover, in both pedigrees, individuals with MM or LGMD2B phenotypes share the same haplotypes. This raises the intriguing possibility that the two diseases may arise from the same gene defect and that a particular disease phenotype is the result of 25 modification by additional factors.

A 3-Mb PAC contig spanning the entire MM/LGMD2B candidate region was recently constructed to facilitate the cloning of the MM/LGMD2B gene(s) (Liu et al., 1998, Genomics 49:23-29). This high resolution PAC contig resolved the discrepancies of the order of markers in previous studies (Bejaoui et al., 1998, Neurogenetics 1:189-96; Bashir et al., 1996, Genomics 33:46-52; Hudson et al., 1995, Science 270:1945-54). The physical size of the PAC contig also indicated that the previous minimal

- 13 -

size estimation based on YAC mapping data was significantly underestimated.

#### Identification of Repeat Sequences and Repeat Typing

The PAC contig spanning the MM/LGMD2B region (Liu et al., 1998, Genomics 49:23-29) was used as a source for the isolation of new informative markers to narrow the genetic interval of the disease gene(s). DNA from the PAC clones spanning the MM/LGMD2B region was spotted onto Hybond N+™ membrane filters (Amersham, Arlington Heights, IL). The filters were hybridized independently with the following γ-³²P (Du Pont, Wilmington, DE) labeled repeat sequences: (1) (CA)<sub>15</sub>; (2) pool of (ATT)<sub>10</sub>, (GATA)<sub>8</sub> and (GGAA)<sub>8</sub>; (3) pool of (GAAT)<sub>8</sub>, (GGAT)<sub>8</sub> and (GTAT)<sub>8</sub>; and (4) pool of (AAG)<sub>10</sub> and (ATC)<sub>10</sub>. Hybridization and washing of the filters were carried out at 55°C following standard protocols (Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual (2nd Edition), Cold Spring Harbor Press, N.Y.).

Miniprep DNAs of PAC clones containing repeat 20 sequences were digested with restriction enzymes HindIII and PstI and ligated into pBluescript II (KS+) vector which is (Stratagene, La Jolla, CA) digested with the same enzymes. Filters of the PAC subclones were hybridized to the  $\gamma$ -32P labeled repeats that detected the 25 respective PACs. For clones with an insert size greater than 1 kb the repeat sequences of which could not be identified by a single round of sequencing, the inserts were further subcloned by digestion with HaeIII and ligation in EcoRV-digested pZero-2.1 vector (Invitrogen, 30 Inc., Carlsbad, CA). Miniprep DNAs of the positive subclones were subjected to manual dideoxy sequencing with Sequenase™ enzyme (US Biochemicals, Inc., Cleveland, OH). Primer pairs for amplifying the repeat sequences were selected using the computer program Oligo (Version

- 14 -

4.0, National Biosciences, Inc., Plymouth, MN). Primer sequences are shown in Table 1.

BNSDOCID: <WO\_\_\_\_0011157A1\_I\_>

TABLE 1

	Het <sup>2</sup>	0.82	0.72	0.30	0.41	0.32
Region	No. of alleles <sup>1</sup>	10	٢	w	색	41
MM/LGMD2B F	Size in PAC (bp)	138	199	161	280	211
fapped to the	Annealing Im (°C)	57	56	ខ	χ 8	95
New Polymorphic Markers Mapped to the MM/LGMD2B Region	Primers (5' to 3')	GATCTAACCCTGCTGCTCACC (SEQ ID NO:120) CTGGTGTGTTGCAGAGCGCTG (SEQ ID NO:121)	CCTCTCTTCTGCTGTCTTCAG (SEQ ID NO:122) TGTGTCTGGTTCCACCTTCGT (SEQ ID NO:123)	TCCAAATAGAAATGCCTGAAC (SEQ ID NO:124) AGGTATCACCTCCAAGTGTTG (SEQ ID NO:125)	TACCAGCTTCAGAGCTCCCTG (SEQ ID NO:126) TTGATCAGGGTGCTCTTGG (SEQ ID NO:127)	GGAGAATTGCTTGAACCCAG (SEQ ID NO:128) TGGCTAATGATGTTGAACATTT (SEQ ID NO:129)
	Repeat	ę.	CCAT	CAT	Complex	AAGG
	Marker	PAC3 - H52	Cy172-H32³	PAC35-PH2	PAC16-H41	Су7-Рнз

Observed in 50 unrelated caucasians. Heterozygosity index. Located within intron 2 of the dysferlin gene.

All oligonucleotides were synthesized by Integrated DNA Technologies, Inc. (Coralville, IA). PCR typing of the repeat markers followed previously described protocols (Bejaoui et al., 1995, Neurology 45:768-772).

- 16 -

Identification of Repeat Markers and Haplotype Analysis After hybridization with labeled repeat oligos, 17 different groups of overlapping PACs were identified that contained repeat sequences. Some groups contained 5 previously identified repeat markers. For example, five groups of PACs were positively identified by a pool of repeat probes including (ATT)<sub>10</sub>, (GATA)<sub>8</sub>, and (GGAA)<sub>8</sub>. Of these, three groups contained known markers GGAA-P7430 (GGAA repeat), D2S1394 (GATA repeat) and D2S1398 (GGAA 10 repeat) (Hudson et al., 1992, Nature 13:622-29; Gastier et al., 1995, Hum. Molecular Genetics 4:1829-36). No attempt was made to isolate new repeat markers from these PACs and they were not further analyzed. Similarly, seven groups of PACs that contained known CA repeat 15 markerswere excluded. Seven groups of PACs that contained unidentified repeats were retained for further analysis. For each group, the PAC containing the smallest insert was selected for subcloning. Subclones were re-screened and positive clones were sequenced to 20 identify repeats. In total, seven new repeat sequences were identified within the MM/LGMD2B PAC contig. Of these, five are polymorphic within the population that was tested. The information for these five markers is summarized in Table 1. Based on the PAC contig 25 constructed previously across the MM candidate locus (Liu et al., 1998, Genomics 48:23-29), the five new markers and ten previously published polymorphic markers were

These markers were analyzed in a large,

30 consanguineous MM family (Bejaoui et al., 1995, Neurology
45: 768-72; Bejaoui et al., 1998, Neurogenetics 1:18996). Because MM is a recessive condition, the locus can
be defined by identifying regions of the genome that show
homozygosity in affected individuals. Conversely,

35 because of the high penetrance of this adult-onset

placed in an unambiguous order (Fig. 1).

- 17 -

condition, unaffected adult individuals are not expected to be homozygous by descent across the region. Analysis of haplotype homozygosity in this pedigree indicates that the disease gene lies between markers D2S2111 and PAC3-H52. Based on the PAC mapping data, the physical distance for this interval is approximately 2.0 Mb. No recombination events were detected between four informative markers (markers cy172-H32 to PAC16-H41) and the disease locus in family MM-21 (Fig. 1A).

#### 10 Identification of Five Muscle-Expressed ESTs

Twenty-two ESTs and two genes (transforming growth factor alpha [TGFα] and beta-adducin [ADD2]) were previously mapped to the MM/LGMD2B PAC contig (Fig. 1A) (Liu et al., 1998, Genomics 48:23-29). Two μl

15 (approximately 0.1 ng/μl) of Marathon-ready™ skeletal muscle cDNA (Clontech, Palo Alto, CA) were used as template in a 10 μl PCR reaction for analysis of muscle expression of ESTs. The PCR conditions were the same as for the PCR typing of repeat markers. PCR analysis of skeletal muscle cDNA indicated that five of these ESTs (A006G04, stsG1553R, WI-14958, TIGR-A004Z44 and WI-14051) map within the minimal genetic MM interval of MM and are expressed in skeletal muscle.

Probes were selected corresponding to each of
these five ESTs for Northern blot analysis. cDNA clones
(130347, 48106, 172575, 184080, and 510138) corresponding
to the five ESTs that are expressed in muscle
(respectively TIGR-A004Z44, WI-14051, WI-14958, stSG1553R
and A006G04) were selected from the UniGene database
(http:/www.ncbi.nlm.nih.gov/UniGene/) and obtained from
Genome Systems, Inc. (St. Louis, MO). The cDNA probes
were first used to screen the MM/LGMD2B PAC filters to
confirm that they mapped to the expected position in the
MM/LGMD2B contig.

- 18 -

A Northern blot (Clontech) of multiple human tissues was sequentially hybridized to the five cDNA probes and a control  $\beta$ -actin cDNA at 65°C following standard hybridization and washing protocols (Sambrook et al., supra). Between hybridizations, probes were removed by boiling the blot at 95-100°C for 4-10 min with 0.5% SDS. The blot was then re-exposed for 24 h to confirm the absence of previous hybridization signals before proceeding with the next round of hybridization.

The tissue distribution, intensity of the signals and size of transcripts detected by the five cDNA probes varied. Probes corresponding to ESTs stSG1553R, TIGR-A004Z44 and WI-14958 detected strong signals in skeletal muscle. In addition, the cDNA corresponding to TIGR-

15 A004Z44 detected a 3.6-3.8 kb brain-specific transcript instead of the 8.5 kb message that was present in other tissues. It is likely that these five ESTs correspond to different genes since the corresponding cDNA probes used for Northern analysis derive from the 3' end of messages,

20 map to different positions in the MM/LGMD2B contig (Fig. 1A), and differ in their expression patterns.

Current database analysis suggests that three of these ESTs (stSG1553R, WI-14958 and WI-14051) do not match any known proteins (Schuler et al., 1996, Science 274:540-46). A006G04 has weak homology with a protein sequence of unknown function that derives from C. elegans. TIGR-A004Z44 has homology only to subdomains present within protein kinase C. Because the five genes corresponding to the ESTs are expressed in skeletal 30 muscle and map within the minimal genetic interval of the MM/LGMD2B gene(s), they are candidate MM/LGMD2B gene(s).

- 19 -

#### Cloning of Dysferlin cDNA

EST TIGR-A004Z44 gave a particularly strong skeletal muscle signal on the Northern blot. Moreover, it is bracketed by genetic markers that show no recombination with the disease phenotype in family MM-21 (Fig. 1). The corresponding transcript was therefore cloned and analyzed as a candidate MM gene. From the Unigene database, a cDNA IMAGE clone (130347, 979 bp) was identified that contained the 483 bp EST TIGR-A004Z44.

Approximately 1 x 10<sup>6</sup> recombinant clones of a λgtl1 human skeletal muscle cDNA library (Clontech) were plated and screened following standard techniques (Sambrook et al., supra). The initial library screening was performed using the insert released from the clone 130347 that

15 contains EST TIGR-A0044Z44, corresponding to the 3' end of the gene. Positive phages were plaque purified and phage DNA was isolated according to standard procedures (Sambrook et al., supra). The inserts of the positive clones were released by EcoRI digestion of phage DNA and subsequently subcloned into the EcoRI site of pBluescript II (KS+) vector (Stratagene).

Fifty cDNA clones were identified when a human skeletal muscle cDNA library was screened with the 130347 cDNA. Clone cDNA10 with the largest insert (~6.5 kb)

25 (Fig. 1B) was digested independently with BamHI and PstI and further subcloned into pBluescript vector. Miniprep DNA of cDNA clones and subclones of cDNA10 was prepared using the Qiagen plasmid Miniprep kit (Valencia, CA). Sequencing was carried out from both ends of each clone using the SequiTherm EXCELTM long-read DNA sequencing kit (Epicenter, Madison, WI), fluorescent-labeled M13 forward and reverse primers, and a LI-COR sequencer (Lincoln, NE). Assembly of cDNA contigs and sequence analysis were performed using Sequencher software (Gene Codes

35 Corporation, Inc., Ann Arbor, MI).

- 20 -

Two additional screens, first with the insert of cDNA10 and then a 683 bp PCR product (A27-F2R2) amplified from the 5' end of the cDNA contig, identified 87 additional cDNA clones. Clones B22 and B33 extended the 5' end by 94 and

20 bp, respectively. The compiled sequence allowed for the generation of a sequence of 6.9 kb (SEQ ID NO:1) (with 10-fold average coverage).

Although the 5' end of the gene has not been further extended to the 8.5 kb predicted by Northern analysis, an open reading frame (ORF) of 6,243 bp has been identified within this 6.9 kb sequence. This ORF is preceded by an in-frame stop codon and begins with the sequence cgcaagcATGCTG (SEQ ID NO:118); five of the first seven bp are consistent with the Kozak consensus sequence for a start codon (Kozak, 1989, Nucl. Acids Res. 15:8125-33; Kozak, 1989, J. Cell. Biol. 108:229-41). An alternate start codon, in the same frame, +75 bp downstream, appears less likely as a start site GAGACGATGGGG (SEQ ID NO:119). Thus, the entire coding region of this candidate gene is believed to have been identified, as represented by the 6.9 kb sequence contig.

# Isolation of the Brain-Specific Dysferlin Isoform Identification of the brain-specific isoform of dysferlin

A brain-specific isoform of dysferlin was identified using Northern blot analysis of poly(A+)RNA derived from multiple human adult tissues probed with radiolabeled full-length dysferlin cDNA subclones. A prominent 7.2 kb transcript was detected on Northern blots in skeletal muscle, heart, placenta, lung, and kidney, while a distinct but equally prominent 3.6 kb-3.8 kb transcript was identified exclusively in the brain. Using long exposures, a faint 7.2 kb mRNA was also detected in the

25

- 21 -

brain. This finding suggested that the shorter brain isoform was likely to be a tissue-specific splice variant of the dysferlin gene. To test this hypothesis, a human brain cDNA library (Stratagene) was screened for the dysferlin brain isoform.

Cloning of the brain-specific dysferlin isoform
To identify probes that hybridize to the brainspecific dysferlin sequence and so could be used for
library screening, fragments of the full-length dysferlin
cDNA clone (derived from a skeletal muscle cDNA library)
were generated using restriction enzymes. The fragments
were about 1 kb in length and were analyzed by
hybridization to a Northern blot that included brain RNA.
Sequences suitable for library screening were those that
hybridized to the 3.6-3.8 kb brain-specific transcript.
A region of the 3' end of the dysferlin cDNA sequence
that is approximately 3 kb in length was identified as
hybridizing to brain mRNA. DNA containing sequence from
this region was used as a probe for hybridization
screening of a human brain cDNA library (Stratagene).

The human brain cDNA library was plated out and screened using standard procedures. Of the approximately 720,000 plaques screened, 63 primary positive clones were identified. Of these, 20 clones were selected for further analysis involving standard methods of hybridization, restriction enzyme mapping, and sequencing. The primary positive clones shared regions of overlap with each other.

Sequencing of positive clones, provided 3671
30 nucleotides of the brain-specific dysferlin sequence (SEQ ID NO:232; Figure 6A-B). The identified sequence corresponds closely to the size of the brain-specific dysferlin transcript detected on Northern blots. With the exception of the 5' region of the sequence, the

- 22 -

brain-specific sequence is identical to about 3.1 kb of the dysferlin sequence (from nucleotide 3722 to 6904 of the dysferlin sequence). In the dysferlin gene, position 3722 corresponds to the start of exon 32. This finding is consistent with the hypothesis that the brain isoform is a splice-variant of the dysferlin gene. At the 5' end of the brain isoform, 489 nucleotides are unique to brain-specific dysferlin. The amino acid sequence encoded by the brain dysferlin nucleic acid sequence (SEQ ID NO:233; Figure 6) contains a unique sequence with an initiation codon within a Kozak consensus sequence. The nucleic acid sequence unique to brain-specific dysferlin encodes a novel 24 amino acid sequence.

#### Identification of Mutations in Miyoshi Myopathy

15 Two strategies were used to determine whether this 6.9 kb cDNA (SEQ ID NO:1) is mutated in MM. First, the genomic organization of the corresponding gene was determined and the adjoining intronic sequence at each of the 55 exons which make up the cDNA was identified. 20 identify exon-intron boundaries within the gene, PAC DNA was extracted with the standard Qiagen -Mini Prep protocol. Direct sequencing was performed with DNA Sequence System (Promega, Madison, WI) using 32P endlabeled primers (Benes et al., 1997, Biotechniques 23:98-25 100). Exon-intron boundaries were identified as the sites where genomic and cDNA sequences diverged. Second, in patients for whom muscle biopsies were available, RT-PCR was also used to prepare cDNA for the candidate gene from the muscle biopsy specimen.

Single strand conformational polymorphism analysis (SSCP) was used to screen each exon in patients from 12 MM families. Putative mutations identified in this way were confirmed by direct sequencing from genomic DNA using exon-specific intronic primers. Approximately 20

- 23 -

ng of total genomic DNA from immortalized lymphocyte cell lines were used as a template for PCR amplification analysis of each exon using primers (below) located in the adjacent introns. SSCP analysis was performed as 5 previously described (Aoki et al., 1998, Ann. Neurol. 43:645-53). In patients for whom muscle biopsies were available, mRNA was isolated using RNA-STAT-60™ (Tel-Test, Friendswood, TX) and first-strand cDNA was synthesized from 1-2  $\mu g$  total RNA with MMLV reverse 10 transcriptase and random hexamer primers (Life Technologies, Gaithersburg, MD). Three  $\mu$ l of this product were used for PCR amplification. Eight sets of primers were designed for muscle cDNA, and overlapping cDNA fragments suitable for SSCP analysis were amplified. 15 After initial denaturation at 94°C for 2 min, amplification was performed using 30 cycles at 94°C for 30 s, 56°C for 30 s, and 72°C for 60 s. The sequences of polymorphisms detected by SSCP analysis were determined

by the dideoxy termination method using the Sequenase kit
(US Biochemicals). In some instances, the base pair
changes predicted corresponding changes in restriction
enzyme recognition sites. Such alterations in
restriction sites were verified by digesting the relevant
PCR products with the appropriate restriction enzymes.

25 Primer pairs used for SSCP screening and exon sequencing are as follows:

- (1) exon 3, F3261 5'-tctcttctcctagagggccatag-3' (SEQ ID NO: 101) and R326 5'-ctgttcctcccatcgtctcatgg-3' (SEQ ID NO: 102);
- 30 (2) exon 20, F3121 5'-gctcctcccgtgaccctctg-3' (SEQ ID NO: 103) and R3121 5'-gggtcccagccaggagcactg-3' (SEQ ID NO: 104);
- (3) exon 36, F2102 5'-cccctctcaccatctcctgatgtg-3'
  (SEQ ID NO: 105) and R2111 5'-tggcttcaccttccctctacctcgg35 3' (SEQ ID NO: 106);

- 24 -

```
(4) exon 49, F1081 5'-tcctttggtaggaaatctaggtgg-3'
   (SEQ ID NO: 107) and R1081 5'-ggaagctggacaggcaagagg-3'
   (SEQ ID NO: 108);
        (5) exon 50, F1091 5'-atatactgtgttggaaatcttaatgag-3'
 5 (SEQ ID NO: 109) and R1091 5'-gctggcaccacagggaatcgg-3'
   (SEQ ID NO: 110);
        (6) exon 51, F1101 5'-ctttgcttccttgcatccttctctg-3'
   (SEQ ID NO: 111) and R1101 5'-agcccccatgtgcagaatggg-3'
   (SEQ ID NO: 112);
        (7) exon 52, F1111 5'-ggcagtgatcgagaaacccgg-3' (SEQ
10
   ID NO: 113) and R1111 5'-catgecetecaetggggetgg-3' (SEQ ID
   NO: 114);
        (8) exon 54, F1141 5'-ggatgcccagttgactccggg-3' (SEQ ID
   NO: 115) and R1141 5'-ccccaccacagtgtcgtcagg-3' (SEQ ID NO:
15 116);
        (9) exon 29, F3031 5'-aagtgccaagcaatgagtgaccgg-3' (SEQ
   ID NO: 184) and R3021 5'-ctcactcccacccaccacctg-3' (SEQ ID
   NO: 185);
        (10) exon 31, F2141 5'-gaatctgccataaccagcttcgtg-3' (SEQ
20 ID NO: 188) and R2141 5'-tatcaccccatagaggcctcgaag-3' (SEQ ID
   NO: 189);
        (11) exon 32, F2981 5'-cagccactcactctggcacctctg-3' (SEQ
   ID NO: 190) and R2981 5'-ageceacagtetetgaetetectg-3' (SEQ ID
   NO: 191);
25
        (12) exon 43, F2031 5'-cagccaaaccatatcaacaatg-3' (SEQ
   ID NO: 210) and R2021 5'-ctggggaggtgagggctctag-3' (SEQ ID
   NO: 211);
        (13) exon 44, F2011 5'-gaagtgttttgtctcctcctc-3' (SEQ ID
   NO: 212) and R2011 5'-gcaggcagccagccccatc-3' (SEQ ID NO:
30 213):
         (14) exon 46, F1041 5'-ctcgtctatgtcttgtgcttgctc-3' (SEQ
   ID NO: 216) and R1051 5'-caccatggtttggggtcatgtqq-3' (SEQ ID
```

NO: 217).

- 25 -

These primers were used in SSCP screening and exon sequencing, and identified eighteen different mutations in fifteen families (Table 2).

BNSDOCID: <WO\_\_\_\_0011157A1\_1\_>

Name	Nucleotide Change	Exon	Consequence	Origin	Family name	Allele	Change of restriction site
Mutations 537insA	ins of A at 537	٣	Frameshift	Arabic	MM59	Нош	no change
Q605X	<u>C</u> AG to <u>T</u> AG at 2186	20	Stop at 605	French	MM67	Нош	-Pst I, -Fnu 4H I¹
I1298V	<u>A</u> TC to <u>G</u> TC at 4265	9 8	Amino acid change	Italian	MM, LGMD56	Het	-BamHI, -BStYI; +Ava II
E1883X	$\underline{\mathtt{G}\mathtt{A}\mathtt{G}}$ to $\underline{\mathtt{I}\mathtt{A}\mathtt{G}}$ at $58,70$	4 9	Stop at 1883	English	MM8	Het	no change
H1857R	C <u>A</u> T to C <u>G</u> T at 5943	20	Amino acid change	English	MM50	Het	no change

no change	no change	no change	no change	-Fnu4HI	-HinPI, -Fsp I	-Mboll	-ScrFI, -BstNI,
Hom	Ном	Het	Het	Het	Hom	Hom	Hom
DMAT71	MM75	MM58	ММ8	MM56	MM10	MM1.7	MM46
Spanish	Spanish	English	English	Italian	Japanese	Japanese	Mexican
Frameshift	Frameshift	Frameshift	5' splice site	Amino acid change	Amino acid change	Frameshift	Stop at 1160
50	20	51	52	54	59	37	32
del of G at 5966	del of G at 5966	del of AG at 6071/6072	<b>Ggt</b> to <u>Ga</u> t at 6319+1	<u>C</u> GT to <u>T</u> GT at 6497	$C\underline{G}C$ to $C\underline{A}G$ at 3510	del of G at 3746	$\underline{C}AG$ to $\underline{\mathtt{I}}AG$ at 3851
5966de1G	5966delG	6071/6072de 1AG	6319+1G to A	R2042C	R1046H	3746delG	Q1160X
			Ŋ				10

5122/5123de de l'CA avin 2000 avin 2	del of CA at 5122/5123, A to T at 5121 CGA to TGA at 5129	4 4 4 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	Frameshift Stop at 1586 Frameshift	Japanese	MM14 MM12	Hom Hom	no change +Dde I
at and cat or at a and and and and and	at 5245 and G to C at 5249, or G to C at 5245 and del G at 5249	<del>1</del> 1					-BanII -BanII, + AvaII, +Sau96I
<u>G</u> AG t at 5	<u>G</u> AG to <u>I</u> AG at 5567	46	Stop at 1732	Spanish	MM73	Нес	II odm-
Ċ.	Del ?Pleas 2573	of ACCCA at e provide -77	CA at 23 ide	Frameshift	ift Italian	an MM69	

ហ

'MM: Miyoshi myopathy; DMAT: distal myopathy with anterior tibial onset; LGMD: limb girdle muscular dystrophy

2 +: create a new restriction site, -: eliminate an existing

restriction site.

- 30 -

Twelve of the eighteen different mutations are predicted to block dysferlin expression, either through nonsense or frameshift changes. Seven of the thirteen samples are homozygous and thus expected to result in complete loss 5 of dysferlin function. For each mutated exon in these patients, at least 50 control DNA samples (100 chromosomes) were screened to determine the frequencies of the sequence variants. When possible, the parents and siblings of affected individuals were also screened to 10 verify that defined mutations were appropriately coinherited with the disease in each pedigree (Fig. 4). two families (50, 58 in Table 2) heterozygous mutations were identified in one allele (respectively a missense mutation and a 2 bp deletion). Mutations in the other 15 allele are presumed to have not been detected (or in three of the screened MM families) either because the mutant and normal SSCP products are indistinguishable or because the mutation lies outside of coding sequence (i.e., in the promoter or a regulatory region of an 20 intron). The disease-associated mutations did not appear to arise in the population as common polymorphisms.

More mutations can be identified by using appropriate primer pairs to amplify an exon and analyze its sequence. The following primer pairs are useful for 25 exon amplification.

	Exon Code		Primer Sequence
	1	F408	5'-gacccacaagcggcgcctcgg-3'{SEQ ID
	NO: 130}		
		F4101	5'-gaccccggcgagggtggtcgg-3'{SEQ ID
30	NO: 131}		
	2	F4111	5'-tgtctctccattctcccttttgtg-3'{SEQ ID
	NO:132}		•
		R4111	5''-aggacactgctgagaaggcacctc-3'{SEQ ID
	NO: 133}		

- 31 -

		3	F3262	5-agtgccctggtggcacgaagg-3' {SEQ ID
	NO:	134}		5 CEO ID
	NO.	135}	R3261	5-cctacctgcaccttcaagccatgg-3' {SEQ ID
5		•	F3251	5-cagaagagccagggtgccttagg-3' {SEQ ID
		136}		
			R3251	5-ccttggaccttaacctggcagagg-3' {SEQ ID
	NO:	137}		- 2/ (GDO TD
10	NIO -	_	F3242	5-cgaggccagcgcaccaacctg-3' {SEQ ID
10	NO:	138}	R3242	5-actgecggecattettgetggg-3' {SEQ ID
	NO:	139}		
		6	F3231	5-ccaggcctcattagggccctc-3' {SEQ ID
	NO:	140}		5
15	NO.	141}	R3231	5-ctgaagaggagcctggggtcag-3' {SEQ ID
	NO.	7	F3222	5-ctgagatttctgactcttggggtg-3' {SEQ ID
	NO:	142}		
			R3211	5-aaggttctgccctcatgccccatg-3' {SEQ ID
20	NO:	143}	70.5.6.1	T The state of the
	NO.	8 144}	F3561	5-ctggcctgagggatcagcagg-3' {SEQ ID
	NO.	144)	R3561	5-gtgcatacatacagcccacggag-3' {SEQ ID
	NO:	145}		
25		9	F3551	5-gagctattgggttggccgtgtggg-3' {SEQ ID
	NO:	146}	22550	F
	NO ·	147}	R3552	5-accaacacggagaagtgagaactg-3' {SEQ ID
	110.	10	F3201	5-ccacactttatttaacgctttggcgg-3'{SEQ
30	ID	NO: 14	18}	
			R3201	5-cagaaccaaaatgcaaggatacgg-3' {SEQ ID
	NO:	149}	<b>72167</b>	
	תד	11 NO: 15	F3191	5-cttctgattctgggatcaccaaagg-3' {SEQ
	110		· • j	

- 32 -

		F3191	5-ggaccgtaaggaagacccaggg-3' {SEQ ID
	NO: 151}		
	12	F3181	5-cctgtgctcaggagcgcatgaagg-3'{SEQ ID
	NO: 152}		
5		R3181	5-gcagacctcccacccaagggcg-3' {SEQ ID
	NO: 153}		
	13	F3171	5-gagacagatgggggacagtcaggg-3' {SEQ ID
	NO: 154}		
		R3171	5-cctcccgagagaaccctcctg-3' {SEQ ID
10	NO: 155}		
	14	F3161	5-gggagcccagagtccccatgg-3' {SEQ ID
	NO: 156}		
	,	R3161	5-gggcctccttgggtttgctgg-3' {SEQ ID
	NO: 157}		333 - 333 - 433 - 433
15	•	F3541	5-gcctccccagcatcctgccgg-3' {SEQ ID
	NO: 158}		
	1.0. 1.00,	R3541	5-tcactgagccgaatgaaactgagg-3' {SEQ
	ID NO: 15		
	16	F3531	5-tgtggcctgagttcctttcctgtg-3' {SEQ ID
20	NO: 160}		
20	1.0. 100,	R3531	5-ggtcaaagggcagaacgaagaggg-3' {SEQ ID
	NO: 161}	10331	·
	17	F3151	5-cccgtccttctcccagccatg-3' {SEQ ID
	NO: 162}	13131	5 deegeeeeeeeeageeaeg-5 (blig 1D
25	NO. 102 j	R3151	5-ctcccctccttctcccc222222/
25	NO. 162)	KSISI	5-ctcccctggttgtccccaagg-3' {SEQ ID
	NO: 163}	E2141	E gangagatatanttaganattata 2/ (GEO ID
	18	F3141	5-cgacccctctgattgccacttgtg-3' {SEQ ID
	NO: 164}	D2141	
		R3141	5-ggcatcctgcccttgccaggg-3' {SEQ ID
30	NO: 165}		
	19	F3522	5-tetgtetecetgeteettg-3' {SEQ ID NO:
	166}		
	_	R3522	5-cttccctgccccgacgcccag-3' {SEQ ID
	NO: 167}		

- 33 -

	20	F3121	5-geteeteegtgaceetetgg-3' {SEQ ID
	NO: 103}	R3121	5-gggtcccagccaggagcactg-3' {SEQ ID
	NO: 104}	K3121	3-gggccccagccaggagcaccg 3 (bbg 1b
5	21	F3111	5-cagcgctcaggcccgtctctc-3' {SEQ ID
	NO: 168}		
		R3111	5-tgcataggcatgtgcagctttggg-3' {SEQ ID
	NO: 169}		
10	22	F3512	5-catgcaccctctgccctgtgg-3' {SEQ ID
10	NO: 170}	R3512	5-agttgagccaggagaggtggg-3' {SEQ ID
	NO: 171}	110022	
	23	F3101	5-catcaggcgcattccatctgtccg-3' {SEQ ID
	NO: 172}		
15	,	R3091	5-agcaggagagcagaagaaagg-3' {SEQ ID
	NO: 173}	F12.0.0.0	F stateton contagging 2/ (SEO ID
	24 NO: 174}	F3082	5-gtgtgtcaccatccccaccccg-3' {SEQ ID
	110. 174	R3082	5-caagagatgggagaaaggccttatg-3' {SEQ
20	ID NO:175	}	
	. 25	F3073	5-ctgggacatccggatcctgaagg-3' {SEQ ID
	NO: 176}		
	NO 177)	R3073	5-tccaggtagtgggaggcagagg-3' {SEQ ID
25	NO: 177}	F3061	5-tcccactacctggagctgccttgg-3' {SEQ
2,5	ID NO: 17	_	5 22242242233343
		R3051	5-ggctctccccagccctccctg-3' {SEQ ID
	NO: 179}		
		F3601	5-cagagcagcagagactctgaccag-3' {SEQ
30	ID NO: 18	·	5 1 (070 77
	NO: 181}	R3601	5-tagaccccacctgcccctgag-3' {SEQ ID
	•	F3501	5-tecteteattgettgeetgttegg-3' {SEQ
	ID NO: 18		J J J J J

- 34 -

	R3501	5-ttgagagcttgccggggatgg-3' {SEQ ID
	NO: 183}	
	29 F3031	5-aagtgccaagcaatgagtgaccgg-3' {SEQ
	ID NO: 184}	
5	R3021	5-ctcactcccacccacctg-3' {SEQ ID
	NO: 185}	
	30 F3011	5-cccaccggcctctgagtctgc-3' {SEQ ID
	NO: 186}	F
	R3001	5-accctacccaagccaggacaagtg-3' {SEQ
10	ID NO: 187}	
	31 F2141	5-gaatetgeeataaceagettegtg-3' {SEQ
	ID NO: 188}	,
	R2141	5-tatcaccccatagaggcctcgaag-3' {SEQ
	ID NO: 189}	
15	32 F2981	5-cagccactcactctggcacctctg-3' {SEQ
	ID NO: 190}	
	R2981	5-agcccacagtctctgactctcctg-3' {SEQ
	ID NO: 191}	
	33 F2131	5-acatctctcagggtccctgctgtg-3' {SEQ
20	ID NO: 192}	
	R2211	5-cctgtgagggacgaggcagg-3' {SEQ ID
	NO: 193}	
	34 F2202	5-gccctgggtaagggatgctgattc-3' {SEQ
	ID NO: 194}	
25	R2202	5-cctgcctgggcctcctggatc-3' {SEQ ID
	NO: 195}	
	35 F2111	5-gagggtgatgggggccttagg-3' {SEQ ID
	NO: 196}	
	R2112	5-gcaatcagtttgaagaaggaaagg-3' {SEQ
30	ID NO: 197}	
	36 F2102	5-cccctctcaccatctcctgatgtg-3' {SEQ
	ID NO: 105}	
	R2111	5-ggcttcaccttccctctacctcgg-3' {SEQ
	ID NO: 106}	

- 35 -

	37	F2101	5-cacetttgtctccattctacctgc-3' {SEQ
	ID NO: 198}		
		R2101	5-ctcccagccccacgcccagg-3' {SEQ ID
	NO: 199}		
5	38	F2091	5-ctgagccactctcctcattctgtg-3' {SEQ
	ID NO: 20	0 }	
		R2091	5-tggaaggggacagtagggagg-3' {SEQ ID
	NO: 201}		
	39	F2081	5-ggccagtgcgttcttcctcctc-3' {SEQ ID
10	NO: 202}		
		R2071	5-tccctgacctgcccatcatctc-3' {SEQ ID
	NO: 203}		
	40	F2061	5-gcccctgtcaggcctggatgg-3' {SEQ ID
	NO: 204}		
15		R2061	5-tgacccaggcctccctggagg-3' {SEQ ID
	NO: 205}		
	41	F2051	5-ctgaaatggtctctttctttctac-3' {SEQ
	ID NO: 20	6}	
		R2051	5-cacaccgactgtcagactgaagag-3' {SEQ
20	ID NO: 20	7}	
	42	F2041	5-ttgtcccctcctctaatccccatg-3' {SEQ
	ID NO: 20	8}	
		R2041	5-gggttagggacgtcttcgagg-3' {SEQ ID
	NO: 209}		
25	43	F2031	5-cagccaaaccatatcaacaatg-3' {SEQ ID
	NO: 210}		
		R2021	5-ctggggaggtgagggctctag-3' {SEQ ID
	NO: 211}		
	44	F2011	5-gaagtgttttgtctcctcctc-3' {SEQ ID
30	NO: 212}		
		R2011	5-gcaggcagccagccccatc-3' {SEQ ID
	NO: 213}		
	45	F1021	5-gggtgccctgtgttggctgac-3' {SEQ ID
	NO: 214}		

- 36 **-**

```
R1031
                       5-gcaggcagccagccccatc-3' {SEQ ID
   NO: 215}
        46
             F1041
                       5-ctcgtctatgtcttgtgcttgctc-3'
                                                       {SEQ
   ID NO: 216}
 5
             R1051
                       5-caccatggtttggggtcatgtgg-3' {SEQ ID
   NO: 217}
        47
             F1061
                        5-tctcgcttccccagctcctgc-3' {SEQ ID
   NO: 218}
                                                   {SEQ ID
             R1061
                       5-tctggagttcgaggactctggg-3'
10 NO: 219}
        48
             F1071
                       5-agaagggtggggagagaacgg-3'
                                                    {SEQ ID
   NO: 220}
                       5-cagctcagagcctgtggctgg-3'
             R1071
   NO: 221}
15
        49
             F1082
                       5-aaggccttcccatcctttggtagg-3'
   ID NO: 222}
             R1082
                       5-acaacccagaggagcacggg-3' {SEQ ID
   NO: 223}
        50
             F1092
                        5-gttgacgatgtatatactgtgttgg-3' {SEQ
20 ID NO: 224}
                        5-gctggcaccacagggaatcgg-3' {SEQ ID
             R1091
   NO: 110}
        51
                        5-gcctctctctaactttgcttccttg-3' {SEQ
             F1102
   ID NO: 225}
25
                        5-agccccatgtgcagaatggg-3' {SEQ ID
             R1101
   NO: 112}
        52
             F1112
                        5-ggctacaggctggcagtgatcgag-3' {SEQ
    ID NO: 226}
                        5-ttccccatgcctccactgg-3' {SEQ ID
             R1112
30 NO: 227}
             F1121
         53
                        5-agccttcgtgcccctaaccaagtg-3'
    ID NO: 228}
                        5-ctgtgggcattggggctcagg-3' {SEQ ID
             R1121
   NO: 229}
```

- 37 -

5-ggatgcccagttgactccggg-3' {SEQ ID 54 F1141 NO: 115} {SEQ ID 5-ccccaccagtgtcgtcagg-3' R1141 NO: 116} {SEQ ID 5-qcccaqtqqqatcaccatg-3' 5 55 F1151 NO: 230} 5-atgctggagggaccccacgg-3' R116 NO: 231}

#### Comparison of Dysferlin With Other Proteins

The 6,243 bp ORF of this candidate MM gene is predicted to encode 2,080 amino acids (Figs. 1C and 2; SEQ ID NO:2). At the amino acid level, this protein is highly homologous to the nematode (Caenorhabditis elegans) protein fer-1 (27% identical, 57% identical or similar: the sequence alignment and comparison was performed using http://vega.igh.cnrs.fr/bin/nph-align\_query.pl.) (Argon & Ward, 1980, Genetics 96:413-33; Achanzar & Ward, 1997, J. Cell Science 110:1073-81). This dystrophy-associated, fer-1-like protein has therefore been designated "dysferlin."

The fer-1 protein was originally identified through molecular genetic analysis of a class of fertilization-defective *C. elegans* mutants in which spermatogenesis is abnormal (Argon & Ward, 1980, *Genetics* 96:413-33). The mutant fer-1 spermatozoa have defective mobility and show imperfect fusion of membranous organelles (Ward et al., 1981, *J. Cell Bio*. 91:26-44). Like fer-1, dysferlin is a large protein with an extensive, highly charged hydrophilic region and a single predicted membrane spanning region at the carboxy terminus (Fig. 3). There is a membrane retention sequence 3' to the membrane spanning stretch, indicating that the protein may be preferentially targeted to either endoplasmic or sarcoplasmic reticulum, probably as a Type II protein

(i.e. with the  $\mathrm{NH_2}$  end and most of the following protein located within the cytoplasm) (Fig. 1C). Several nuclear membrane targeting sequences are predicted within the cytoplasmic domain of the protein

5 (http://psort.nibb.ac.jp/form.html). Immunocytochemical detection of dysferlin suggests that dysferlin is targeted to or anchored within the sarcoplasmic reticulum.

The cytoplasmic component of this protein contains

10 four motifs homologous to C2 domains. C2 domains are
intracellular protein modules composed of 80 - 130 amino
acids (Rizo & Sudhof, 1998, J. Biol. Chem. 273:15897).

Originally identified within a calcium-dependent isoform
of protein kinase C (Nishizuka, 1988, Nature 334:661-65),

- 15 C2 domains are present in numerous proteins. These domains often arise in approximately homologous pairs described as double C2 or DOC2 domains. One DOC2 protein, DOC2α, is brain specific and highly concentrated in synaptic vesicles (Orita et al., 1995, Biochem.
- 20 Biophys. Res. Comm. 206:439-48), while another, DOC2β, is ubiquitously expressed (Sakaguchi et al., 1995, Biochem. Biophys. Res. Comm. 217:1053-61). Many C2 modules can fold to bind calcium, thereby initiating signaling events such as phospholipid binding. At distal nerve
- 25 terminals, for example, the synaptic vesicle protein synaptotagmin has two C2 domains that, upon binding calcium, permit this protein to interact with syntaxin, triggering vesicle fusion with the distal membrane and neurotransmitter release (Sudhof & Rizo, 1996, Neuron 17:379-88).

The four dysferlin C2 domains are located at amino acid positions 32-82, 431-475, 1160-1241, and 1582-1660 (Figs. 1C and 3). Indeed, it is almost exclusively through these regions that dysferlin has homology to any 35 proteins other than fer-1. Each of these segments in

- 39 -

dysferlin is considerably smaller than a typical C2
domain. Moreover, these segments are more widely
separated in comparison with the paired C2 regions in
synaptotagmin, DOC2α and β and related C2-positive
5 proteins. For this reason, it is difficult to predict
whether the four relatively short C2 domains in dysferlin
function analogously to conventional C2 modules. That
dysferlin might, by analogy with synaptotagmin, signal
events such as membrane fusion is suggested by the fact
10 that fer-1 deficient worms show defective membrane
organelle fusion within spermatozoa (Ward et al., 1981,
J. Cell Bio. 91:26-44).

The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

#### **EXAMPLES**

### Example 1: Production of dysferlin protein

Standard methods can be used to synthesize either wild type or mutant dysferlin, or fragments of either.

- These methods can also be used to synthesize brainspecific dysferlin polypeptides including full-length or
  fragments (e.g., a polypeptide unique to brain-specific
  dysferlin). For example, a recombinant expression vector
  encoding dysferlin (or a fragment thereof: e.g.,
- 25 dysferlin minus its membrane-spanning region) operably linked to appropriate expression control sequences can be used to express dysferlin in a prokaryotic (e.g., E.coli) or eukaryotic host (e.g., insect cells, yeast cells, or mammalian cells). The protein is then purified by
- 30 standard techniques. If desired, DNA encoding part or all of the dysferlin sequence can be joined in-frame to DNA encoding a different polypeptide, to produce a chimeric DNA that encodes a hybrid polypeptide. This can be used, for example, to add a tag that will simplify
- 35 identification or purification of the expressed protein,

- 40 -

or to render the dysferlin (or fragment thereof) more immunogenic.

The preferred means for making short peptide fragments of dysferlin is by chemical synthesis. These fragments, like dysferlin itself, can be used to generate antibodies, or as positive controls for antibody-based assays.

Fusion proteins are useful, e.g., for generating antibodies. Such fusion proteins are generated using 10 known methods. In one example, to construct glutathione S-transferase (GST): dysferlin fusion proteins, the BLAST program (Altschul et al., 1990, J. Molec. Biol. 215:403-410) was used to identify three regions of the dysferlin cDNA that show no homology to any known human proteins 15 (Figure 1). These were subcloned from the dysferlin cDNA as BstYI (881-1333), XmnI (1990-2718) and SalI (5364-5732) fragments ligated respectively into BamHI, SmaI and SalI sites of pGEX-5X-3 (Pharmacia). The three fragments correspond to amino acid sequences at amino acid 20 locations 253-403, 624-865, and 1664-1786 of SEQ ID NO:2, respectively. The resulting GST fusion proteins of BamHI (43 kDa) and SmaI (53.3 kDa) formed isoluble aggregates that were isolated by SDS-PAGE. The fusion protein of SalI (40.2 kDa) was soluble and thus could be purified 25 using a glutathione Sepharose 4B column; the SalI dysferlin fragment (14.2 kDa) was isolated by cleavage from GST using Factor Xa protease. The eluted protein was concentrated and further purified by SDS-PAGE. all three of the fusion peptides, the resulting SDS-PAGE 30 bands were excised and used to immunize rabbits.

# Example 2: Production and characterization of antidysferlin antibodies

Techniques for generating both monoclonal and polyclonal antibodies specific for a particular protein

- 41 -

are well known. The antibodies can be raised against a short peptide epitope of dysferlin, an epitope linked to a known immunogen to enhance immunogenicity, a long fragment of dysferlin, or the intact protein. Antibodies can also be raised against brain-specific dysferlin polypeptides, e.g., against amino acids 1-24 of SEQ ID NO:233. Such antibodies raised against dysferlin or brain-specific dysferlin polypeptides are useful for e.g., localizing such polypeptides in tissue sections or fractionated cell preparations and diagnosing dysferlin-related disorders.

An isolated dysferlin protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind dysferlin using standard techniques 15 for polyclonal and monoclonal antibody preparation. dysferlin immunogen can also be a mutant dysferlin or a fragment of a mutant dysferlin. A full-length dysferlin protein can be used or, alternatively, antigenic peptide fragments of dysferlin can be used as immunogens. 20 antigenic peptide of dysferlin comprises at least 8 (preferably 10, 15, 20, or 30) amino acid residues of the amino acid sequence shown in SEQ ID NO:2 and encompasses an epitope of such that an antibody raised against the peptide forms a specific immune complex with dysferlin. 25 Preferred epitopes encompassed by the antigenic peptide are regions of dysferlin that are located on the surface of the protein, e.g., hydrophilic regions.

A dysferlin immunogen typically is used to prepare antibodies by immunizing a suitable subject (e.g., 30 rabbit, goat, mouse or other mammal) with the immunogen. An appropriate immunogenic preparation can contain, for example, recombinantly expressed dysferlin protein or a chemically synthesized dysferlin polypeptide. The preparation can further include an adjuvant, such as 35 Freund's complete or incomplete adjuvant, or similar

- 42 -

immunostimulatory agent. Immunization of a suitable subject with an immunogenic dysferlin preparation induces a polyclonal anti-dysferlin antibody response.

Polyclonal anti-dysferlin antibodies ("dysferlin antibodies") can be prepared as described above by immunizing a suitable subject with a dysferlin immunogen. The dysferlin antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using

- immobilized dysferlin. If desired, the antibody molecules directed against dysferlin can be isolated from the mammal (e.g., from the blood) and further purified by well-known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after
- immunization, e.g., when the dysferlin antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein
- 20 (1975) Nature 256:495-497, the human B cell hybridoma technique (Kozbor et al. (1983) Immunol. Today 4:72), the EBV-hybridoma technique (Cole et al. (1985), Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for
- producing hybridomas is well known (see generally Current Protocols in Immunology (1994) Coligan et al. (eds.) John Wiley & Sons, Inc., New York, NY). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with a
- 30 dysferlin immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds dysferlin.

Any of the many well known protocols used for fusing 35 lymphocytes and immortalized cell lines can be applied

- 43 -

for the purpose of generating a monoclonal antibody against dysferlin (see, e.g., Current Protocols in Immunology, supra; Galfre et al. (1977) Nature 266:55052; R.H. Kenneth, in Monoclonal Antibodies: A New Dimension 5 In Biological Analyses, Plenum Publishing Corp., New York, New York (1980); and Lerner (1981) Yale J. Biol. Med., 54:387-402. Moreover, the one in the art will appreciate that there are many variations of such methods which also would be useful. Hybridoma cells producing a 10 monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind dysferlin, e.g., using a standard ELISA assay.

Alternative to preparing monoclonal antibody-15 secreting hybridomas, a monoclonal dysferlin antibody can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (e.g., an antibody phage display library) with dysferlin to thereby isolate immunoglobulin library members that bind dysferlin. Kits 20 for generating and screening phage display libraries are commercially available (e.g., the Pharmacia Recombinant Phage Antibody System, Catalog No. 27-9400-01; and the Stratagene SurfZAP™ Phage Display Kit, Catalog No. 240612). Additionally, examples of methods and reagents 25 particularly amenable for use in generating and screening antibody display library can be found in, for example, U.S. Patent No. 5,223,409; PCT Publication No. WO 92/18619; PCT Publication No. WO 91/17271; PCT Publication No. WO 92/20791; PCT Publication No. WO 30 92/15679; PCT Publication No. WO 93/01288; PCT Publication No. WO 92/01047; PCT Publication No. WO 92/09690; PCT Publication No. WO 90/02809; Fuchs et al. (1991) Bio/Technology 9:1370-1372; Hay et al. (1992) Hum. Antibod. Hybridomas 3:81-85; Huse et al. (1989) Science

- 44 -

246:1275-1281; Griffiths et al. (1993) *EMBO J*. 12:725-734.

As an example, two polyclonal antisera were raised for each of the fusion peptide antigens described above 5 using New Zealand White rabbits. The rabbits were injected with 0.5 mg of antigen using keyhole limpet hemocyanin (KLH) as the adjuvent. Booster injections of 0.25 mg antigen were administered every three weeks over 12 weeks. Serum was prepared from the rabbits and was purified using affinity column chromatography (HiTrap; Pharmacia) or antigen-blotted polyvinylidene difluoride (PVDF) membrane.

Immunoblotting was used to verify that the affinitypurified antisera recognize the cognate fusion peptides

15 by Western immunoblotting (WIB) and that this reactivity
was immunoadsorbed by pre-incubation of the antisera with
the peptides. Thus, antiserum raised against the
polypeptide encoded by the SalI fragment (encoding amino
acids 1664-1786) identified the fragment both as a

20 cleaved, 14.2 kDa fragment and as a component of the 40.2
kDa GST-SalI fusion peptide. No reactivity was evident
in the fraction containing only the GST fusion partner.
Immunoadsorption entirely abolished this staining.
Analogous results were detected with all six antisera (to
25 the three different target fusion peptides).

#### Preparation of subcellular fractions

Frozen human muscle (0.3 g) was homogenized in five volumes of 0.25 M sucrose containing proteinase inhibitor (Complete, Boehringer). Subcellular fractions of nuclei, 30 mitochondria, microsomes, and cytosol were separated by differential centrifugation. The purity of each fraction was evaluated by immunoblotting of fraction-specific proteins with antibodies to histone H1 (Calbiochem), cytochrome c (Santa Cruz), Na\*-K\* ATPase α1 subunit

- 45 -

(Research Diagnostics) and cytosolic superoxide dismutase (Calbiochem).

#### Dysferlin in subcellular fractions

Immunoblotting was used to analyze dysferlin 5 expression. Twenty  $\mu g$  of each subcellular fraction and 40  $\mu$ q of whole homogenate of muscle were separated by SDS-PAGE (4-15% gradient gel) and transferred to a nitrocellulose membrane. Immunoblotting was performed according to standard methods, using chemiluminescence Immunoblotting of multi-tissue blots 10 (ECL, Amersham). identified prominent dysferlin positively at approximately 230 kDa in heart, placenta, skeletal muscle and kidney. Little or no immuno-positive staining was detected in brain, liver, spleen, ovary, or testis. 15 Lower molecular weight bands (approximately 40 kDa) were also evident. Immunoadsorption with the corresponding fusion peptide abolished both the large and the smaller bands. The 230 kDa band was observed with all of the affinity purified, anti-dysferlin antisera.

Immunoblotting of fractionated human muscle documented distinct 230 kDa bands in the whole muscle homogenate an in microsomal and nuclear fractions. Some immunoreactivity was also evident in the nuclear and mitochondrial fractions. No immunoreactivity was detected in the cytosolic fractions. This pattern was seen with all of the anti-dysferlin antisera, and was eliminated by immunoadsorption. The identity of the assayed fractions was verified by Western blotting using fraction-specific antibodies: histone HI for the nuclear fraction, cytochrome c for the mitochondrial fraction, Na\*-K\* ATPase α1-subunit for the microsomal fraction, and SOD1 for the cytosolic fraction.

### Example 3: Diagnosis

The discovery of mutations in the dysferlin gene that are associated with the MM and LMGD2B phenotypes means that individuals can be tested for the disease gene before symptoms appear. This will permit genetic testing 5 and counseling of those with a family history of the disease. Additionally, individuals diagnosed with the genetic defect can be closely monitored for the appearance of symptoms, thereby permitting early intervention, including genetic therapy, as appropriate. 10 Individuals with a brain-specific dysferlin-related

disorder can be diagnosed using such methods.

Diagnosis can be carried out on any suitable genomic DNA sample from the individual to be tested. Typically, a blood sample from an adult or child, or a sample of 15 placental or umbilical cord cells of a newborn would be used; alternatively, one could utilize a fetal sample obtained by amniocentesis or chorionic villi sampling.

It is expected that standard genetic diagnostic methods can be used. For example, PCR can be utilized to 20 identify the presence of a deletion, addition, or substitution of one or more nucleotides within any one of the exons of dysferlin. Following the PCR reaction, the PCR product can be analyzed by methods such as a heteroduplex detection technique based upon that of White 25 et al. (1992, *Genomics* 12:301-06), or by techniques such as cleavage of RNA-DNA hybrids using RNase A (Myers et al., 1985, Science 230:1242-46), single-stranded conformation polymorphism (SSCP) analysis (Orita et al., 1989, Genomics 10:298-99), di-deoxy-fingerprinting (DDF) 30 (Blaszyk et al., 1995, Biotechniques 18: 256-260) and denaturing gradient gel electrophoresis (DGGE; Myers et al., 1987, Methods Enzymol. 155:501-27). The PCR may be carried out using a primer which adds a G+C rich sequence (termed a "GC-clamp") to one end of the PCR product, thus 35 improving the sensitivity of the subsequent DGGE

- 47 -

procedure (Sheffield et al., 1989, Proc. Natl. Acad. Sci. USA 86:232-36). If the particular mutation present in the patient's family is known to have removed or added a restriction site, or to have significantly increased or decreased the length of a particular restriction fragment, a protocol based upon restriction fragment length polymorphism (RFLP) analysis (perhaps combined with PCR) may be appropriate.

The apparent genetic heterogeneity resulting in the

10 MM/LGMD2B phenotypes means that the nature of the

particular mutation carried by affected individuals in

the patient's family may have to be ascertained prior to

attempting genetic diagnosis of the patient.

Alternatively, a battery of tests designed to identify

15 any of several mutations known to result in MM/LGMD2B may

be utilized to screen individuals without a defined

familial genotype. The analysis can be carried out on

any genomic DNA derived from the patient, typically from

a blood sample.

Instead of basing the diagnosis on analysis of the 20 genomic DNA of a patient, one could seek evidence of the mutation in the level or nature of the relevant expression products. Well-known techniques for analyzing expression include mRNA-based methods, such as Northern 25 blots and in situ hybridization (using a nucleic acid probe derived from the relevant cDNA), and quantitative PCR (as described in St-Jacques et al., 1994, Endocrinology 134:2645-57). One could also employ polypeptide based methods, including the use of 30 antibodies specific for the polypeptide of interest. These techniques permit quantitation of the amount of expression of a given gene in the tissue of interest, at least relative to positive and negative controls. One would expect an individual who is heterozygous for a 35 genetic defect affecting the level of expression of

dysferlin to show up to a 50% loss of expression of this gene in such a hybridization or antibody-based assay. An antibody specific for the carboxy terminal end would be likely to pick up (by failure to bind to) most or all 5 frameshift and premature termination signal mutations, as well as deletions of the carboxy terminal sequence. of a battery of monoclonal antibodies specific for different epitopes of dysferlin would be useful for rapidly screening cells to detect those expressing mutant 10 forms of dysferlin (i.e., cells which bind to some dysferlin-specific monoclonal antibodies, but not to others), or for quantifying the level of dysferlin on the surface of cells. One could also use a protein truncation assay (Heim et al., 1994, Nature Genetics 15 8:218-19) to screen for any genetic defect which results in the production of a truncated polypeptide instead of the wild type protein.

Use of immunodetection to identify normal and disease-associated dysferlin

In the following example, immunodetection methods are used to demonstrate a detectable difference in muscles homogenates between normal and disease-associated dysferlin alleles.

Frozen muscle samples (quadriceps) were homogenized
in ten volumes of SDS-PAGE sample buffer and boiled for 5
minutes. The final loading volume of SDS-PAGE was
adjusted after densitometric measurements (NIH Image) of
myosin heavy chain on the Coomassie blue stained gels.
Studies were performed on six MM, two LGMD-2B, and three
normal muscle samples.

Immunocytochemistry was performed on 8 micron cryostat sections of the muscle that were fixed in 100% cold acetone for 5 minutes and preincubated with PBS containing 1% BSA, 5% heat-inactivated goat serum and 35 0.2% Triton®X-100. The sections were incubated with

- 49 -

primary antibodies overnight at 4°C and fluorescein-labeled secondary (TAGO Immunologicals) for 30 minutes at room temperature. The primary antibodies were applied in two double staining combinations: SalI-1 anti-dysferlin and anti-dystrophin antibodies, and SalI-2 anti-dysferlin and anti- $\delta$ -sarcoglycan antibodies. The sections were mounted in SlowFade (Molecular Probes).

The 230 kDA antigen was absent in samples from all five MM patient in immunoblot assays. All five patients 10 had normal patterns of dystrophin expression. Genetic analysis of the dysferlin gene in the patients predicted that at least two of the five MM patients should have no full-length protein. Two of the other three patients had mutations in at least one allele that are predicted to eliminate normal dysferlin expression. In all five patients, absence of dysferlin immuno-staining was documented with at least two other anti-dysferlin anti-sera.

Immunostaining of dysferlin, dystrophin and  $\delta$ 20 sarcoglycan proteins demonstrated distinct membraneassociated positivity for each protein in normal muscle.
By contrast, in both MM and LGMD-2B muscle the dysferlin
protein was absent, while the dystrophin and  $\delta$ sarcoglycan proteins appeared normal.

## 25 Therapeutic Treatment

A patient with MM/LGMD2B, or an individual genetically susceptible to contracting one or both of these diseases, can be treated by supplying dysferlin therapeutic agents of the present invention. Dysferlin therapeutic agents include a DNA or a subgenomic polynucleotide coding for a functional dysferlin protein. A DNA (e.g., a cDNA) is prepared which encodes the wild type form of the gene operably linked to expression control elements (e.g., promoter and enhancer) that

- 50 -

induce expression in skeletal muscle cells or any other affected cells. The DNA may be incorporated into a vector appropriate for transforming the cells, such as a retrovirus, adenovirus, or adeno-associated virus. 5 of the many other known types of techniques for introducing DNA into cells in vivo may be used (e.g., liposomes). Particularly useful would be naked DNA techniques, since naked DNA is known to be readily taken up by skeletal muscle cells upon injection into muscle. 10 Wildtype dysferlin protein can also be administered to an individual who either expresses mutant dysferlin protein

or expresses an inadequate amount of dysferlin protein, e.g., a MM/LGMD2B patient.

Administration of the dysferlin therapeutic agents 15 of the invention can include local or systemic administration, including injection, oral administration, particle gun, or catheterized administration, and topical administration. Various methods can be used to administer the therapeutic dysferlin composition directly 20 to a specific site in the body. For example, a specific muscle can be located and the therapeutic dysferlin composition injected several times in several different locations within the body of the muscle. therapeutic dysferlin composition can be directly 25 administered to the surface of the muscle, for example, by topical application of the composition. X-ray imaging can be used to assist in certain of the above delivery methods. Combination therapeutic agents, including a dysferlin protein or polypeptide or a subgenomic 30 dysferlin polynucleotide and other therapeutic agents, can be administered simultaneously or sequentially.

Receptor-mediated targeted delivery of therapeutic compositions containing dysferlin subgenomic polynucleotides to specific tissues can also be used. 35 Receptor-mediated DNA delivery techniques are described

- 51 -

in, for example, Findeis et al. (1993), Trends in Biotechnol. 11, 202-05; Chiou et al. (1994), Gene Therapeutics: Methods and Applications of Direct Gene Transfer (J.A. Wolff, ed.); Wu & Wu (1988), J. Biol.
5 Chem. 263, 621-24; Wu et al. (1994), J. Biol. Chem. 269, 542-46; Zenke et al. (1990), Proc. Natl. Acad. Sci. U.S.A. 87, 3655-59; Wu et al. (1991), J. Biol. Chem. 266, 338-42.

Alternatively, a dysferlin therapeutic composition

10 can be introduced into human cells ex vivo, and the cells then implanted into the human. Cells can be removed from a variety of locations including, for example, from a selected muscle. The removed cells can then be contacted with the dysferlin therapeutic composition utilizing any of the above-described techniques, followed by the return of the cells to the human, preferably to or within the vicinity of a muscle. The above-described methods can additionally comprise the steps of depleting fibroblasts or other contaminating non-muscle cells subsequent to removing muscle cells from a human.

Both the dose of the dysferlin composition and the means of administration can be determined based on the specific qualities of the therapeutic composition, the condition, age, and weight of the patient, the 25 progression of the disease, and other relevant factors. If the composition contains dysferlin protein or polypeptide, effective dosages of the composition are in the range of about 1  $\mu$ g to about 100 mg/kg of patient body weight, e.g., about 50  $\mu$ g to about 50 mg/kg of patient body weight, e.g., about 500  $\mu$ g to about 5 mg/kg of patient body weight.

Therapeutic compositions containing dysferlin subgenomic polynucleotides can be administered in a range of about 0.1  $\mu g$  to about 10 mg of DNA/dose for local administration in a gene therapy protocol. Concentration

- 52 -

ranges of about 0.1  $\mu$ g to about 10 mg, e.g., about 1  $\mu$ g to about 1 mg, e.g., about 10  $\mu$ g to about 100  $\mu$ g of DNA can also be used during a gene therapy protocol. such as method of action and efficacy of transformation 5 and expression are considerations that will effect the dosage required for ultimate efficacy of the dysferlin subgenomic polynucleotides. Where greater expression is desired over a larger area of tissue, larger amounts of dysferlin subgenomic polynucleotides or the same amounts 10 readministered in a successive protocol of administrations, or several administrations to different adjacent or close tissue portions of for example, a muscle site, may be required to effect a positive therapeutic outcome. In all cases, routine 15 experimentation in clinical trials will determine specific ranges for optimal therapeutic effect.

#### Animal Model

A line of transgenic animals (e.g., mice, rats, guinea pigs, hamsters, rabbits, or other mammals) can be 20 produced bearing a transgene encoding a defective form of dysferlin. Standard methods of generating such transgenic animals would be used, e.g., as described below.

Alternatively, standard methods of producing null

(i.e., knockout) mice could be used to generate a mouse which bears one defective and one wild type allele encoding dysferlin. If desired, two such heterozygous mice could be crossed to produce offspring which are homozygous for the mutant allele. The homozygous mutant offspring would be expected to have a phenotype comparable to the human MM and/or LGMD2B phenotype, and so serve as models for the human disease.

For example, in one embodiment, dysferlin mutations are introduced into a dysferlin gene of a cell, e.q., a

- 53 -

fertilized oocyte or an embryonic stem cell. Such cells can then be used to create non-human transgenic animals in which exogenous altered (e.g., mutated) dysferlin sequences have been introduced into their genome or 5 homologously recombinant animals in which endogenous dysferlin nucleic acid sequences have been altered. Such animals are useful for studying the function and/or activity of dysferlin and for identifying and/or evaluating modulators of dysferlin function. As used 10 herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, 15 dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene 20 product in one or more cell types or tissues of the transgenic animal. As used herein, an "homologously recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous dysferlin gene has been altered by homologous 25 recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to completed development of the animal.

A transgenic animal of the invention can be created

30 by introducing a nucleic acid encoding a dysferlin
 mutation into the male pronuclei of a fertilized oocyte,
 e.g., by microinjection or retroviral infection, and
 allowing the oocyte to develop in a pseudopregnant female
 foster animal. A dysferlin cDNA sequence e.g., that of

35 (SEQ ID NO:1 or SEQ ID NO:3) can be introduced as a

- 54 -

transgene into the genome of a non-human animal. Alternatively, a nonhuman homologue of the human dysferlin gene can be isolated based on hybridization to the human dysferlin sequence (e.g., cDNA) and used as a 5 transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, 10 have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, U.S. Patent No. 4,873,191 and in Hogan, Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). 15 Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the mutant dysferlin transgene in its genome and/or expression of the mutant dysferlin mRNA in tissues or cells of the

20 animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene.

Moreover, transgenic animals carrying a transgene encoding a mutant dysferlin can further be bred to other transgenic animals carrying other transgenes.

To create an homologously recombinant animal, a vector is prepared which contains at least a portion of a dysferlin gene into which a deletion, addition or substitution has been introduced to thereby alter a dysferlin gene. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous dysferlin gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous dysferlin gene is mutated

- 55 -

or otherwise altered (e.g., contains one of the mutations described in Table 2). In the homologous recombination vector, the altered portion of the dysferlin sequence is flanked at its 5' and 3' ends by additional nucleic acid 5 of the dysferlin gene to allow for homologous recombination to occur between the exogenous dysferlin nucleic acid sequence carried by the vector and an endogenous dysferlin gene in an embryonic stem cell. additional flanking dysferlin nucleic acid is of 10 sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see, e.g., Thomas and Capecchi (1987) Cell 51:503 for a description of homologous recombination 15 vectors). The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced dysferlin sequence has homologously recombined with the endogenous dysferlin gene are selected (see, e.g., Li et al. (1992) Cell 69:915). 20 selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras (see, e.g., Bradley in Teratocarcinomas and Embryonic Stem Cells: A Practical Approach, Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be 25 implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline 30 transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley (1991) Current Opinion in Bio/Technology 2:823-829 and in PCT Publication Nos. WO 90/11354, WO 91/01140, WO 35 92/0968, and WO 93/04169.

- 56 -

## Other Embodiments

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

BNSDOCID: <WO\_\_\_\_\_0011157A1\_I\_>

- 57 **-**

What is claimed is:

- 1. An isolated DNA comprising a nucleotide sequence which hybridizes under stringent hybridization conditions to SEQ ID NO:3, or a complement thereof.
- 5 2. The isolated DNA of claim 1, wherein the nucleotide sequence is SEQ ID NO:117.
  - 3. An isolated DNA comprising a nucleotide sequence selected from the group consisting of SEQ ID NOs:4-12.
- 4. The isolated DNA of claim 3, comprising the sequence of SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, or SEQ ID NO:21.
  - 5. An isolated DNA comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:22-30.
- 15 6. A single stranded oligonucleotide of 14-50 nucleotides in length having a nucleotide sequence identical to a portion of SEQ ID NO:3, or a complement thereof.
  - 7. A pair of PCR primers consisting of:
- 20 (a) a first single stranded oligonucleotide consisting of 14-50 contiguous nucleotides that are identical to a portion of SEQ ID NO:117; and
- (b) a second single stranded oligonucleotide consisting of 14-50 contiguous nucleotides that are 25 identical to a portion of SEQ ID NO:117, wherein the sequence of at least one of the oligonucleotides is identical to a portion of a strand of SEQ ID NO:3, and the first oligonucleotide is not complementary to the second oligonucleotide.

- 58 -

- 8. A pair of single-stranded oligonucleotides, wherein both oligonucleotides are selected from the group consisting of SEQ ID NOS:130-231, SEQ ID NO:110, and SEQ ID NO:112 and the oligonucleotides are different from 5 each other.
  - 9. An isolated DNA comprising a nucleotide sequence that encodes a polypeptide that shares at least 70% sequence identity with SEQ ID NO:2, or a complement of the nucleotide sequence.
- 10. The isolated DNA of claim 9, wherein the polypeptide comprises the sequence of SEQ ID NO:2.
- 11. An isolated DNA comprising a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid having a sequence selected from the group consisting of SEQ ID NOs:31-79 and 90-100.
- 12. A single stranded oligonucleotide of 14-50 nucleotides in length comprising a nucleotide sequence which is identical to a portion of a nucleic acid selected from the group consisting of SEQ ID NOs:31-79 and 90-100, or a complement of the nucleotide sequence.
  - 13. The oligonucleotide of claim 12, wherein the portion includes an intronic sequence.
    - 14. A pair of PCR primers consisting of:
- (a) a first single-stranded oligonucleotide
  25 consisting of 14-50 contiguous nucleotides that are identical to a portion of a sense strand of a nucleic acid selected from the group consisting of SEQ ID NOs:31-85; and

PCT/US99/19395

WO 00/11157

- 59 -

- (b) a second single stranded oligonucleotide consisting of 14-50 contiguous nucleotides that are identical to a portion of the antisense strand of a nucleic acid selected from the group consisting of SEQ ID NOs:31-85, wherein the sequence of at least one of the oligonucleotides comprises a sequence identical to a portion of a nucleic acid selected from SEQ ID NOs: 31-79 and 90-100, and wherein the first oligonucleotide is not complementary to the second oligonucleotide.
- 15. A pair of single-stranded oligonucleotides selected from the group consisting of SEQ ID NOs:101-116, SEQ ID NOs:184-185, SEQ ID NOs:188-191, SEQ ID NOs:210-213, and SEQ ID NOs:216-217.
- 16. A vector comprising the isolated DNA of claim15 1.
  - 17. A substantially pure polypeptide comprising an amino acid sequence sharing at least 70% sequence identity with SEQ ID NO:2.
- 18. The substantially pure polypeptide of claim 17, 20 wherein the polypeptide comprises an amino acid sequence identical to that of a naturally occurring polypeptide.
  - 19. The substantially pure polypeptide of claim 18, wherein the amino acid sequence comprises the sequence of SEQ ID NO:2.
- 25 20. A substantially pure polypeptide comprising an amino acid sequence identical to the amino acid sequence of amino acid residues 1-500, 501-1000, 1001-1500, or 1501-2080 of SEQ ID NO:2.

- 60 -

21. A substantially pure polypeptide comprising the amino acid sequence of SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88 or SEQ ID NO:89.

- 22. A substantially pure polypeptide selected from 5 the group consisting of amino acids 253-403 of SEQ ID NO:2, amino acids 624-865 of SEQ ID NO:2, and amino acids 1664-1786 of SEQ ID NO:2.
  - 23. A fusion protein comprising a polypeptide of claim 22.
- 10 24. An antibody that specifically binds to the polypeptide of claim 22.
  - 25. An antibody that binds specifically to the polypeptide of claim 17.
    - 26. A cell comprising the isolated DNA of claim 1.
- 15 27. A non-human mammal, the genomic DNA of which bears a transgene, wherein the transgene comprises the isolated DNA of claim 1.
- 28. A transgenic non-human mammal having a transgene disrupting or interfering with the expression 20 of a dysferlin gene.
  - 29. A method of decreasing the symptoms of muscular dystrophy in a mammal, the method comprising introducing into a cell of said mammal the isolated DNA of claim 1.
- 30. A method of decreasing the symptoms of muscular 25 dystrophy in a mammal, the method comprising introducing

- 61 -

into a cell of said mammal the vector of claim 16, the vector being an expression vector.

- 31. A method of decreasing the symptoms of muscular dystrophy in a mammal, the method comprising introducing 5 into a cell of said mammal the protein of claim 17.
  - 32. A method for identifying a patient, a fetus, or a pre-embryo at risk for having a dysferlin-related disorder, the method comprising:
- (a) obtaining a sample of genomic DNA from the 10 patient, fetus, or pre-embryo; and
  - (b) determining whether the sample contains a mutation in a dysferlin gene, wherein a patient, a fetus, or a pre-embryo having a mutation in a dysferlin gene is at risk for having a dysferlin-related disorder.
- 15 33. The method of claim 32, comprising:
  - (a) treating the sample of genomic DNA with a restriction enzyme specific for a particular restriction enzyme site; and
- (b) detecting the presence or absence of the 20 particular restriction enzyme site in the sample of genomic DNA as an indication of the presence or absence of a particular mutation in the genomic DNA.
- 34. The method of claim 33, wherein the restriction enzyme is selected from the group consisting of Pst I,
  25 Fnu4H I, BamH I, BstY I, Ava II, HinP I, Fsp I, Mbo II,
  ScrF I, BstN I, Mae I, Bfa I, Dde I, Bpm I, Ban II, Ava
  II, and Sau96 I.
  - 35. The method of claim 32, comprising subjecting the sample to polymerase chain reaction (PCR).

- 36. The method of claim 32, comprising:
- (a) contacting a single stranded oligonucleotide with the sample of genomic DNA; and
- (c) detecting hybridization or lack thereof between 5 the single stranded oligonucleotide and the genomic DNA, as an indication of the presence or absence of a mutation in the genomic DNA.
- 37. A method for identifying a patient, a fetus, or a pre-embryo at risk for having a dysferlin-related10 disorder, said method comprising:
  - (a) providing a sample comprising dysferlin mRNA from the patient, fetus, or pre-embryo; and
- (b) determining whether the dysferlin mRNA contains a mutation, wherein a patient, a fetus, or a pre-embryo15 having a dysferlin mRNA containing a mutation is at risk for having a dysferlin-related disorder.
  - 38. The method of claim 37, wherein the presence or absence of the mutation is detected by Northern blot.
- 39. The method of claim 37, wherein the method 20 includes the step of subjecting the sample to polymerase chain reaction (PCR).
  - 40. A method for detecting the absence of a mutation in a dysferlin protein of a patient, a fetus, or a pre-embryo, the method comprising:
- 25 (a) providing a sample comprising a dysferlin protein of the patient, fetus, or pre-embryo;
  - (b) contacting the sample with the antibody of claim 22; and
- (c) detecting binding of the antibody to dysferlin 30 protein in the sample, if any, wherein binding indicates a normal dysferlin protein.

- 63 -

- 41. An isolated DNA comprising a nucleotide sequence that is identical to the sequence of amino acid residues 3501-3520 of SEQ ID NO:1, 3737-3756 of SEQ ID NO:1, 3842-3861 of SEQ ID NO:1, 5114-5139 of SEQ ID NO:1, or 5239-5 5255 of SEQ ID NO:1.
  - 42. An isolated DNA comprising a nucleotide sequence selected from the group consisting of 3501-3520 of SEQ ID NO:1, wherein nucleotide G at 3510 is A;
- 3737-3756 of SEQ ID NO:1, wherein nucleotide G at 3746 is deleted;
  - 3842-3861 of SEQ ID NO:1, wherein nucleotide C at 3851 is T;
    - 5114-5139 of SEQ ID NO:1, wherein nucleotide C at
- 15 5122 and nucleotide A at 5123 are deleted;
  5239-5255 of SEQ ID NO:1, wherein nucleotide G at
  5245 is deleted and nucleotide G at 5249 is C; and
  5239-5255 of SEQ ID NO:1, wherein nucleotide G at
  - 5245 is C and nucleotide G at 5249 is deleted.
- 20 43. An isolated nucleic acid comprising a nucleotide sequence which hybridizes under stringent hybridization conditions to nucleic acids 3284-3720 of SEQ ID NO:232, or the complement of said nucleotide sequence.
- 25 44. An isolated nucleic acid comprising a nucleotide sequence identical to the sequence of nucleotides 3284-3720 of SEQ ID NO:232, or a complement of said nucleotide sequence.
- 45. The isolated nucleic acid of claim 44, wherein 30 the nucleotide sequence comprises the sequence of SEQ ID NO:232 or the complement of SEQ ID NO:232.

- 46. An isolated polypeptide comprising:
- a) at least 15 contiguous amino acids of the polypeptide comprising amino acids 1-24 of SEQ ID NO:233,
- b) a naturally occuring allelic variant of a 5 polypeptide comprising amino acids 1-24 of SEQ ID NO:233, or
  - c) an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes under stringent conditions to nucleotides 3284-3720 of SEQ ID NO:232.
- 10 47. The polypeptide of claim 46, wherein the polypeptide comprises SEQ ID NO:233.
  - 48. A vector comprising the nucleic acid of claim 44.
    - 49. A cell comprising the vector of claim 48.
- 15 50. A method of making a polypeptide, the method comprising culturing the cell of claim 49.
  - 51. An antibody which specifically binds to a polypeptide of claim 46.
- 52. The antibody of claim 51, wherein the antibody 20 binds to a polypeptide selected from the group comprising amino acids 253-403 of SEQ ID NO:233, amino acids 624-865 of SEQ ID NO:233, and amino acids 1664-1786 of SEQ ID NO:233.
- 53. The antibody of claim 51, wherein the antibody 25 is a monclonal antibody.
  - 54. The antibody of claim 51, wherein the antibody is a polyclonal antibody.

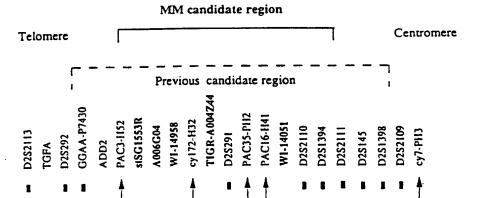


FIG. 1A

FIG. 1B

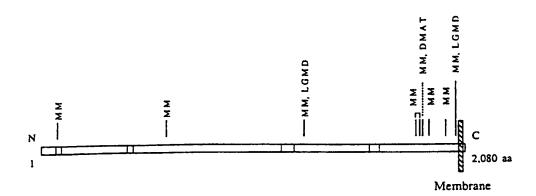
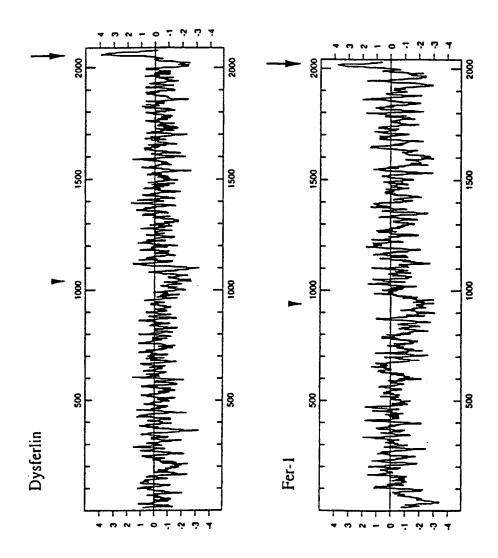


FIG. 1 C

1 MERVEILYAE NVHTPDTDIS DAYCSAVFAG VKKRTKVIKN SVNPVWNEGF 51 EWDLKGIPLD OGSELHVVVK DHETMGRNRF LGEAKVPLRE VLATPSLSAS 101 FNAPLLDTKK QPTGASLVLQ VSYTPLPGAV PLFPPPTPLE PSPTLPDLDV 151 VADTGGEEDT EDOGLTGDEA EPFLDOSGGP GAPTTPRKLP SRPPPHYPGI 201 KRKRSAPTSR KLLSDKPQDF QIRVQVIEGR QLPGVNIKPV VKVTAAGQTK 251 RTRIHKGNSP LFNETLFFNL FDSPGELFDE PIFITVVDSR SLRTDALLGE 301 FRMDVGTIYR EPRHAYLRKW LLLSDPDDFS AGARGYLKTS LCVLGPGDEA 351 PLERKDPSED KEDIESNLLR PTGVALRGAH FCLKVFRAED LPQMDDAVMD 401 NVKQIFGFES NKKNLVDPFV EVSFAGKMLC SKILEKTANP OWNONITLPA 451 MFPSMCEKMR IRIIDWDRLT HNDIVATTYL SMSKISAPGG EIEEEPAGAV 501 KPSKASDLDD YLGFLPTFGP CYINLYGSPR EFTGFPDPYT ELNTGKGEGV 551 AYRGRLLLSL ETKLVEHSEQ KVEDLPADDI LRVEKYLRRR KYSLFAAFYS 601 ATMLQDVDDA IQFEVSIGNY GNKFDMTCLP LASTTQYSRA VFDGCHYYYL 651 PWGNVKPVVV LSSYWEDISH RIETONQLLG IADRLEAGLE QVHLALKAQC 701 STEDVDSLVA QLTDELIAGC SQPLGDIHET PSATHLDQYL YQLRTHHLSQ 751 ITEAALALKI GHSELPAALE QAEDWLLRLR ALAEEPQNSL PDIVIWMLQG 801 DKRVAYORVP AHOVLFSRRG ANYCGKNCGK LQTIFLKYPM EKVPGARMPV 851 QIRVKLWFGL SVDEKEFNQF AEGKLSVFAE TYENETKLAL VGNWGTTGLT 901 YPKFSDVTGK IKLPKDSFRP SAGWTWAGDW FVCPEKTLLH DMDAGHLSFV 951 EEVFENQTRL PGGQWIYMSD NYTDVNGEKV LPKDDIECPL GWKWEDEEWS 1001 TDLNRAVDEQ GWEYSITIPP ERKPKHWVPA EKMYYTERRR RWVRLRRRDL 1051 SOMEALKRER QAFAEGEGWE YASLFGWKFH LEYRKTDAFR RRRWRRRMEP 1101 LEKTGPAAVF ALEGALGGVM DDKSEDSMSV STLSFGVNRP TISCIFDYGN 1151 RYHLRCYMYQ ARDLAAMDKD SFSDPYAIVS FLHOSOKTVV VKNTLNPTWD 1201 OTLIFYEIEI FGEPATVAEO PPSIVVELYD HDTYGADEFM GRCICOPSLE 1251 RMPRLAWFPL TRGSQPSGEL LASFELIQRE KPAIHHIPGF EVQETSRILD 1301 ESEDTDLPYP PPQREANIYM VPQNIKPALQ RTAIEILAWG LRNMKSYQLA 1351 NISSPSLVVE CGGQTVQSCV IRNLRKNPNF DICTLFMEVM LPREELYCPP 1401 ITVKVIDNRQ FGRRPVVGQC TIRSLESFLC DPYSAESPSP QGGPDDVSLL 1451 SPGEDVLIDI DDKEPLIPIQ EEEFIDWWSK FFASIGEREK CGSYLEKDFD 1501 TLKVYDTQLE NVEAFEGLSD FCNTFKLYRG KTQEETEDPS VIGEFKGLFK 1551 IYPLPEDPAI PMPPRQFHQL AAQGPQECLV RIYIVRAFGL OPKDPNGKCD 1601 PYIKISIGKK SVSDODNYIP CTLEPVFGKM FELTCTLPLE KDLKITLYDY 1651 DLLSKDEKIG ETVVDLENRL LSKFGARCGL PQTYCVSGPN QWRDQLRPSQ 1701 LLHLFCQQHR VKAPVYRTDR VMFQDKEYSI EEIEAGRIPN PHLGPVEERL 1751 ALHVLQQQGL VPEHVESRPL YSPLQPDIEQ GKLQMWVDLF PKALGRPGPP 1801 FNITPRRARR EFLRCIIWNT RDVILDDLSL TGEKMSDIYV KGWMIGFEEH 1851 KOKTDVHYRS LGGEGNFNWR FIFPFDYLPA EQVCTIAKKD AFWRLDKTES 1901 KIPARVVFQI WDNDKFSFDD FLGSLQLDLN RMPKPAKTAK KCSLDQLDDA 1951 FHPEWFVSLF EQKTVKGWWP CVAEEGEKKI LAGKLEMTLE IVAESEHEER 2001 PAGQGRDEPN MNPKLEDPRR PDTSFLWFTS PYKTMKFILW RRFRWAIILF 2051 IILFILLEL AIFIYAFPNY AAMKLOKPES (SEQ 10 NO 12)

FIG. 2



Hydrophobicity Index

FIG. 3

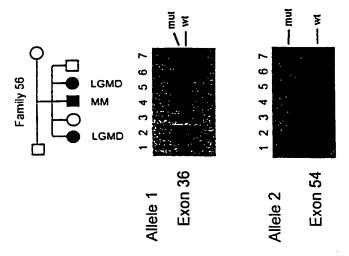
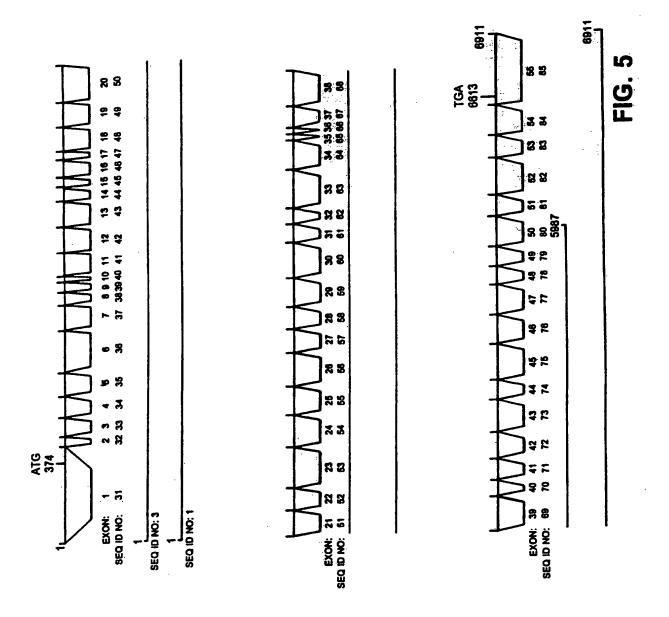


FIG. 4



61/21 TCC TGG TTC AAG CGA TTC TCT GGC CTC AGC CTC CCG AGT AGC TGG GAT TAC AGG CAT GCT CCA CCA AGC CCG GGT AAT TTT GTA TTT TTA S W F K R F S G L S L P S S W D Y R H A P P S P G N F V F L 91/31 151/51 91/31 ATA GAG ACG GGG TIT TGC CAT GIT GGT CAG GCT GGT CTC GAA CTC CTG ACC TCA GGT GAT CTG CCC ACC TTG GCC TCC CAA CGT GCT GAG A G L E L L T S G D L P T L A S Q R A E 211/71 241/81 ETGFCHVGQA 181/61 ATT ACA GGC ATG AGT CAC TGT GCC CGG CAG AGA TGG TCT AAT TCA TAT GAA AGA ACT CTG AAA AAA GTA GAA AGT GAT TTT CTA AAA TAA
I T G M S H C A R Q R W S N S Y E R T L K K V E S D F L K \* 331/111 301/101 271/91 271/91
GGT ACA AAT AAT TAA TGT AAG CAT AAT CAC CTA ACC TTG TGG AAT TTT TTT TTT TTG AGA AGC AAA TTG CAA ATT TGT GAT AGA TCT AAA
G T N N " C K H N H L T L W N F F F L R S K L Q I C D R S K ии • скнинг 391/131 421/141 361/121 GGA GAT TGA CTA AGA GGG TGA CCA TCT GGA AAT GAC GTC ATG TGA GAA TGG TTA AAG ATG CTC GGG AGA TTG AGC CTA GAG AAA GGA AGA G D \* L R G \* 451/151 P S G N D V M \* E W L K M L G R L S L E
481/161 511/171 481/161 451/151
TITT GTG AAC CCA GGA GGC AGA GGT AGA GAT CCA GGA GAG ggc ggc gtg atg gat gac aag agt gaa gat tee atg tee gtc tee acc ttg
F V N P G G R G R D P G E G G V M D D K S E D S M S V S T L

541/181
571/191
601/201 age tte ggt gtg aac aga ccc acg att tcc tgc ata tte gac tat ggg aac cgc tac cat cta cgc tgc tac atg tac cag gcc cgg gac

S F G V N R P T I S C I F D Y G N R Y H L R C Y M Y Q A R D

631/211 661/221 691/231 S F G 631/211 ctg gct gcg atg gac aag gac tct ttt tct gat ccc tat gcc atc gtc tcc ttc ctg cac cag agc cag aag acg gtg gtg gtg aag aac L A A M D K D S F S D P Y A I V S F L H Q S Q K T V V V K N 721/241 751/251 781/261 acc ctt aac ccc acc tgg gac cag acg ctc atc ttc tac gag atc gag atc ttt ggc gag ccg gcc aca gtt gct gag caa ccg ccc agc
T L N P T W D Q T L I F Y E I E I F G E P A T V A E Q P P S
811/271 841/281 871/291 811/271 att gtg gtg gag ctg tac gac cat gac act tat ggt gca gac gag ttt atg ggt cgc tgc atc tgt caa ccg agt ctg gaa cgg atg cca I V V E L Y D H D T Y G A D E F M G R C I C Q P S L E R M P 901/301 931/311 961/321 egg ctg gcc tgg ttc cca ctg acg agg ggc agc cag ccg tcg ggg gag ctg ctg gcc tct ttt gag ctc atc cag aga gag aag ccg gcc R L A W F P L T R G S Q P S G E L L A S F E L I Q R E K P A 1021/341 1051/351 atc cac cat att cet ggt tit gag gtg cag gag aca tca agg atc ctg gat gag tct gag gac aca gac ctg ccc tac cca ccc cag I H H I P G F E V Q E T S R I L D E S E D T D L P Y P P P Q 1111/371 1141/381 1081/361 agg gag gcc aac atc tac atg gtt cct cag aac atc aag cca gcg ctc cag cgt acc gcc atc gag atc ctg gca tgg ggc ctg cgg aac R E A N I Y M V P Q N I K P A L Q R T A I E I L A W G L R N 1231/411 1201/401 1171/391 atg aag agt tac cag ctg gcc aac atc tcc ccc agc ctc gtg gta gag tgt ggg ggc cag acg gtg cag tcc tgt gtc atc agg aac M K S Y Q L A N I S S P S L V V E C G G Q T V Q S C V I R N 1321/441 1291/431 1261/421 ctc cgg aag aac ccc aac ttt gac atc tgc acc ctc ttc atg gaa gtg atg ctg ccc agg gag gag ctc tac tgc ccc ccc atc acc gtc
L R K N P N F D I C T L F M E V M L P R E E L Y C P P I T V

1351/451 1381/461 1411/471 aag gtc atc gat aac cgc cag ttt ggc cgc cgg cct gtg gtg ggc cag tgt acc atc cgc tcc ctg gag agc ttc ctg tgt gac ccc tac K V I D N R Q F G R R P V V G Q C T I R S L E S F L C D P Y 1471/491 1501/501 1441/481 teg geg gag agt eca tee cea cag ggt gge eca gac gat gtg age eta ete agt eet ggg gaa gac gtg ete ate gac att gat gac aag S A E S P S P Q G G P D D V S L L S P G E D V L I D I D D K 1561/521 1591/531 gag coe ete ate cee ate cag gag gaa gag tte ate gat tgg tgg age ama tte ttt gee tee ata ggg gag agg gam ama tge gge tee E P L I P I Q E E E F I D W W S K F F A S I G E R E K C G S 1531/511 1681/561 1651/551 1621/541 tac ctg gag aag gat ttt gac acc ctg aag gtc tat gac aca cag ctg gag aat gtg gag gcc ttt gag ggc ctg tct gac ttt tgt aac Y L E K D F D T L K V Y D T Q L E N V E A F E G L S D F C N 1771/591 1741/581 1711/571 K L 1861/621 1831/611 ctc cca gaa gac cca gcc atc ccc atg ccc cca aga cag ttc cac cag ctg gcc gcc cag gga ccc cag gag tgc ttg gtc cgt atc tac L P E D P A I P M P P R Q F H Q L A A Q G P Q E C L V R I Y 1951/651 1921/641 1891/631 att gtc ega gca ttt ggc ctg cag ccc aag gac ccc aat gga aag tgt gat cct tac atc aag atc tcc ata ggg aag aaa tca gtg agt I V R A F G L Q P K D P N G K C D P Y I K I S I G K K S V S 2041/681 2011/671 1981/661 gac cag gat aac tac atc ccc tgc acg ctg gag ccc gta ttt gga aag atg ttc gag ctg acc tgc act ctg cct ctg gag aag gac cta
D Q D N Y I P C T L E P V F G K M F E L T C T L P L E K D L
2071/691 2101/701 2131/711 aag atc act cte tat gac tat gac ctc ctc tcc aag gac gaa aag atc ggt gag acg gtc gcc ctg gag aac agg ctg ctg tcc aag K I T L Y D Y D L L S K D E K I G E T V V D L E N R L L S K 2161/721 2221/741 2281/761 2311/771 2251/751 2251/51
ctc ttc tgc cag cag cat aga gtc aag gca cct gtg tac cgg aca gac cgt gta atg ttt cag gat aaa gaa tat tcc att gaa gag ata
L F C Q Q H R V K A P V Y R T D R V M F Q D K E Y S I E E I CQQHRVKAPV 2371/791 2401/801 2341/781 gag gct ggc agg atc cca aac cca cac ctg ggc cca gtg gag cgt ctg gct ctg gct ctg cat gtg ctt cag cag cag cag ctg gtc ccg gag E A G R I P N P H L G P V E E R L A L H V L Q Q Q G L V P E

Figure 6A

								2461	/821									2491	l								
2431/811 cac gtg gag				ctc	tac'	1	ccc	cta	Cag	CCA	gac	atc	gag	cag	999	aag	ctg	cag	چ	tgg	gtc	gac	cta	ttt	ccg	aag	gcc
	ECA	R	D	T.	v	S	P	L	0	P	Ď	I	E	Q	G	K	L	Q	M	W	V	D	L	F	P	K	A
H V E									1001									2581	1/861	L							
2521/841 ctg ggg cgg	cct	CCA	cct	ccc	ttc	aac	atc	acc	cca	cgg	aga	gcc	aga	agg	ttt	ttc	ctg	cgt	tgt	att	atc	tgg	aat	acc	aga	gat	grg
L G R	D	G	P	P	F	N	I	T	P	R	R	A	R	R	F	F	L	••	_	-	I	W	N	T	ĸ	ט	٧
								2641	/881									2671	1/891	L						~~	
2611/871 atc ctg gat	GAC	cta	age	ctc	acg	999	gag	aag	atg	agc	gac	att	tat	gtg	aaa	ggt	tgg	atg	att	ggc	EEE	gaa	gaa	Cac	aay v	^	K
I L D	D		s	L	T	G	E	K	M	S	D	I	Y	V	K	G	W	••	-	-	F	E	E	п	•	~	••
								2731	/911	L								2761	1/921					act	gag	CAA	atc
2701/901 aca gac gtg	cat	tat	cgt	tcc	ctg	gga	ggt	gaa	ggc	aac	ttc	aac	tgg	agg	ttc	att	ttc	CCC P	EEC	gac	v	t	D	A	E	0	v
		Y			L	G	G	E	G	N	F	N	W	R	F	1	r	-	./951	_	1		•	••	_	•	
								2821	/941									2001		ata	ata	ttc	CAC	atc	taa	gac	aat
2791/931 tgt acc att	gcc	aag	aag	gat	gcc	ttc	tgg	agg	ctg	gac	aag	act	gag	agc	aaa v	T	D	yca A	D D	y cy	v	P	0	I	W	Ď	N
CTI	A	K	K	D	A	F	W	ĸ	_	ט	V.	7.	E	3	~	-	•	••	7981	•	•	•	-	_			
2881/961 gac aag tto								2911	./9/1				ctc		cac	aro	ccc	880	CCA	acc	aag	aca	qcc	aag	aag	tgc	tcc
gac aag ttc	tcc	ttt	gat	gat	ttt	ctg	ggc	S	ctg	Cay	t.	n	t.	N	R	M	P	K	P	A	K	T	Ã	X .	K	С	S
D K F		F						2001	1100	11								3031	/101	11							
2971/991 ttg gac cag			_					3001	, 100 FAA	**	ata	tcc	ctt	ttt	gag	CAG	aaa	aca	ata	aag	ggc	tgg	tgg	CCC	tgt	gta	gca
	ctg	gat	gat	gct	F	Н	b	E	W	F	v	s	L	F	E	0	K	T	v	K	G	W	W	P	C	V	A
L D Q		D			-		-	3001	1100							_		3121	/104	11							
3061/1021 gaa gag ggt	~			878	cta	aca	aac	aaq	ctg	gaa	atg	acc	ttg	gag	att	gta	gca	gag	agt	gag	cat	gag	gag	cgg	cct	gct	ggc
	gay	z v	K	ī	L	A	G	K	L _	Ē	M	T	L	E	I	V	A	_	-	_	н	E	E	R	P	A	G
E E G																		3211	/107	/1							
3151/1051 cag ggc cgg	CAT	gag	ccc	aac	atq	aac	cct	aag	ctt	gag	gac	CC8	agg	cgc	CCC	gac	acc	tcc	ttc	ctg	tgg	ttt	acc	tcc	CCA	tac	aay v
O G R	D .	E	P	N	м	N	P	K		E	D	P	R	R	P	D	T	-	-	_	W	F	Т	5	r	ı	
								3271	/109	1								3301	/110	)1					~~~	atc	ttc
3241/1081 acc atg aag	ttc	atc	ctg	tgg	cgg	cgt	ttc	cgg	tgg	gcc	atc	atc	ctc	ttc	atc	atc	CEC	F	atc	CEG	L	ctg	F	t.	A	ī	F
	F		L	W	R	R	F	R	W	Λ	1	1	L	F.	1	1	L		/113	_	_	_	•	~		-	•
3331/1111								3361	/112	1							ctc	2221	rac	cct	ota	GAA	aaa	acc	ata	999	tcc
3331/1111 atc tac gcc	ttc	ccg	aac	tat	gct	gcc	atg	aag	ctg	gcg	<b>20</b> 9	D	E	c	tya t	gya	T.	S	C	P	v	E	G	A	v	G	S
atc tac gcc																	_	3481	/116	51							
3421/1141 cct cca gca						+	+	3431	CAG	ctc	aac	gag	ctc	ctc	cag	acc	tcc	tag	gcc	tga	ttg	tcc	tgc	cag	ggt	<b>ggg</b>	cag
		gac	rgg	CCL	A	c	S	A	0	L	G	Ė	L	L	Q	T .	S	•	Ä	•	L	5	С	Q	G	G	Q
P P A		D		-				25 41	7110	11								3571	/119	1							
3511/1171 aca gac aga		200	CCC	CCB	CAC	tcc	cag	agt	tgc	taa	cat	gga	gct	ctg	aga	tca	CCC	cac	ttc	cat	cat	ttc	ctt	ctc	CCC	CAB	ccc
	W			P	н	S	0	S	c	•	н	G	A	L	R	S	P	••	•	••	H	F	L	L	P	Q	P
T D R	-	-					_	2521	/121	4								3661	/122	21				- T	nΩ -	222	
3601/1201 aac gct ttt	tta	GAT.	ca¤	ctc	aga	cat	att	tca	gta	taa	aac	agt	tgg	aac	cac	aaa	aaa	888	200	ممم	aa	ÇSI	W.	ו עם	MU:	<b>Z</b> 3Z	
ade yet tet																											
n a f	L	Ď	Q	L	R	H	I	s	v	•	N	S	W	N	Н	K	K	K	K	K		(SI	2Q 7	ED 1	NO:	233	• •

Figure 6B

## SEQUENCE LISTING

<110> The Ger	neral Hospital	Corporation		
<120> DYSFERI GIRDLE	LIN, A GENE MUT MUSCULAR DYSTR	CATED IN DISTAL ROPHY	MYOPATHY AND LIMB	
<130> 00786/3	399WO2	•		
<150> US 60/0 <151> 1998-08				
<160> 233				
<170> FastSE	Q for Windows V	Version 3.0		
<210> 1 <211> 6911 <212> DNA <213> Homo sa	apiens			
<220> <221> CDS <222> (374).	(6613)			
<pre>&lt;400&gt; 1 tcgaccgccc agccagg agattacagc tcgacgg tgttctcgga acgccgg gcccactgga gcagccg agccagagat tcgagcc ggcgcctcgg ccctccc acacgcgcca agc atg Met</pre>	agc tcgggaaggg ctg acaagcgggg ggg gtggcccgtt ggc ctcgcccagc gac ctttccgagc ctg agg gtc tt Leu Arg Val Pl	cggcgggggt gga tgagcgcagg cgg cccctttaag agc cagccctctc cag cctctttgcg ccc tc atc ctc tat	agatgag cagaagcccc ggcgggg acccagccta aactgct ctaagccagg cgagggg acccacaagc tgggcgc acggggccct	60 120 180 240 300 360 409
cac aca ccc gac ac His Thr Pro Asp Th 15	c gac atc agc o r Asp Ile Ser 1 20	gat gcc tac tgc Asp Ala Tyr Cys	tcc gcg gtg ttt Ser Ala Val Phe 25	457
gca ggg gtg aag aa Ala Gly Val Lys Ly 30	g aga acc aaa q s Arg Thr Lys \ 35	gtc atc aag aac Val Ile Lys Asn 40	Ser Val Asn Pro	505
gta tgg aat gag gg Val Trp Asn Glu Gl 45	a ttt gaa tgg q y Phe Glu Trp i 50	gac ctc aag ggc Asp Leu Lys Gly 55	atc ccc ctg gac Ile Pro Leu Asp 60	553
cag ggc tct gag ct Gln Gly Ser Glu Le 6	t cat gtg gtg o u His Val Val v 5	gtc aaa gac cat Val Lys Asp His 70	gag acg atg ggg Glu Thr Met Gly 75	601
agg aac agg ttc ct Arg Asn Arg Phe Le 80	g ggg gaa gcc u Gly Glu Ala	aag gtc cca ctc Lys Val Pro Leu 85	cga gag gtc ctc Arg Glu Val Leu 90	649
gcc acc cct agt ct Ala Thr Pro Ser Le 95	g tcc gcc agc : u Ser Ala Ser : 100	ttc aat gcc ccc Phe Asn Ala Pro	ctg ctg gac acc Leu Leu Asp Thr 105	697
aag aag cag ccc ac Lys Lys Gln Pro Th 110	a ggg gcc tcg r Gly Ala Ser 115	ctg gtc ctg cag Leu Val Leu Gln 120	Val Ser Tyr Thr	745

ccg Pro 125	ctg Leu	cct Pro	gga Gly	gct Ala	gtg Val 130	ccc Pro	ctg Leu	ttc Phe	ccg Pro	ccc Pro 135	cct Pro	act Thr	cct Pro	ctg Leu	gag Glu 140	793
ccc Pro	tcc Ser	ccg Pro	act Thr	ctg Leu 145	cct Pro	gac Asp	ctg Leu	gat Asp	gta Val 150	gtg Val	gca Ala	gac Asp	aca Thr	gga Gly 155	gga Gly	841
gag Glu	gaa Glu	gac Asp	aca Thr 160	gag Glu	gac Asp	cag Gln	gga Gly	ctc Leu 165	act Thr	gga Gly	gat Asp	gag Glu	gcg Ala 170	gag Glu	cca Pro	889
ttc Phe	ctg Leu	gat Asp 175	caa Gln	agc Ser	gga Gly	ggc Gly	ccg Pro 180	ggg Gly	gct Ala	ccc Pro	acc Thr	acc Thr 185	cca Pro	agg Arg	aaa Lys	937
cta Leu	cct Pro 190	tca Ser	cgt Arg	cct Pro	ccg Pro	ccc Pro 195	cac His	tac Tyr	ccc Pro	gly ggg	atc Ile 200	aaa Lys	aga Arg	aag Lys	cga Arg	985
agt Ser 205	gcg Ala	cct Pro	aca Thr	tct Ser	aga Arg 210	aag Lys	ctg Leu	ctg Leu	tca Ser	gac Asp 215	aaa Lys	ccg Pro	cag Gln	gat Asp	ttc Phe 220	1033
cag Gln	atc Ile	agg Arg	gtc Val	cag Gln 225	gtg Val	atc Ile	gag Glu	ggg Gly	cgc Arg 230	cag Gln	ctg Leu	ccg Pro	ggg Gly	gtg Val 235	aac Asn	1081
atc Ile	aag Lys	cct Pro	gtg Val 240	gtc Val	aag Lys	gtt Val	acc Thr	gct Ala 245	gca Ala	<b>G</b> JÀ GGG	cag Gln	acc Thr	aag Lys 250	cgg Arg	acg Thr	1129
cgg <b>A</b> rg	atc Ile	cac His 255	Lys	gga Gly	aac Asn	agc Ser	cca Pro 260	Leu	ttc Phe	aat Asn	gag Glu	act Thr 265	Dea	ttc Phe	ttc Phe	1177
aac Asn	ttg Leu 270	Phe	gac Asp	tct Ser	cct Pro	ggg Gly 275	gag Glu	ctg Leu	ttt Phe	gat Asp	gag Glu 280	FLO	atc Ile	ttt Phe	atc Ile	1225
acg Thr 285	· Val	gta Val	. gac . Asp	tct Ser	cgt Arg 290	Ser	ctc Leu	agg Arg	aca Thr	gat Asp 295	мта	ctc	ctc Leu	ggg Gly	gag Glu 300	1273
tto Phe	cgg Arg	ato Met	. Asp	, Val	ggc Gly	Thr	TIE	Tyr	ALU	GIU	ccc Pro	cgg	cac His	gcc Ala 315	tat Tyr	1321
cto	agg Arg	aaq Lys	tgg Trp 320	) Leu	ctg Leu	ctc Leu	tca Ser	gac Asp 325	PIC	gat Asp	gac Asp	tto Phe	tct Ser 330		ggg	1369
gco	aga a Arg	ggq g Gly 33!	, Tyr	cto Lev	g aaa Lys	aca Thr	ago Ser 340	. Let	tgt Cys	gto Val	g ctg Lev	g ggg Gly 345	FIC	ggç Gly	gac Asp	1417
gaa Gl	a gcg a Ala 350	Pro	cto Le	g gaq ı Glı	g aga 1 Arg	aaa g Lys 355	ASI	c ccc p Pro	tct Ser	gaa Glu	a gad 1 Asp 360	י עם י	g gaç s Glu	g gad 1 Asp	att Ile	1465
ga: G1: 36:	u Sei	c aa c As	c cto n Leo	g cto 1 Leo	c cgg 1 Arg 370	y Pro	aca Thi	a ggo r Gly	gta Y Val	a gco l Ala 37!	a nec	g ega 1 Arq	a gga g Gly	a gco 7 Ala	cac His 380	1513
tt Ph	c tgo e Cyr	c ct	g aaq u Ly:	g gte s Val 38	l Phe	c cgg	g gce g Ala	c gaq a Gli	g gad u Asj 39	b re	g ccq u Pro	g caq o Gli	g ato n Med	g gad Asj 39	gat p Asp	1561

gcc Ala	gtg Val	atg Met	gac Asp 400	aac Asn	gtg Val	aaa Lys	cag Gln	atc Ile 405	ttt Phe	ggc Gly	ttc Phe	gag Glu	agt Ser 410	aac Asn	aag Lys	1609
aag Lys	aac Asn	ttg Leu 415	gtg Val	gac Asp	ccc Pro	ttt Phe	gtg Val 420	gag Glu	gtc Val	agc Ser	ttt Phe	gcg Ala 425	ggg Gly	aaa Lys	atg Met	1657
ctg Leu	tgc Cys 430	agc Ser	aag Lys	atc Ile	ttg Leu	gag Glu 435	aag Lys	acg Thr	gcc Ala	aac Asn	cct Pro 440	cag Gln	tgg Trp	aac Asn	cag Gln	1705
aac Asn 445	atc Ile	aca Thr	ctg Leu	cct Pro	gcc Ala 450	atg Met	ttt Phe	ccc Pro	tcc Ser	atg Met 455	tgc Cys	gaa Glu	aaa Lys	atg Met	agg Arg 460	1753
att Ile	cgt Arg	atc Ile	ata Ile	gac Asp 465	tgg Trp	gac Asp	cgc Arg	ctg Leu	act Thr 470	cac His	aat Asn	gac Asp	atc Ile	gtg Val 475	gct Ala	1801
acc Thr	acc Thr	tac Tyr	ctg Leu 480	agt Ser	atg Met	tcg Ser	aaa Lys	atc Ile 485	tct Ser	gcc Ala	cct Pro	gga Gly	gga Gly 490	gaa Glu	ata Ile	1849
gaa Glu	gag Glu	gag Glu 495	cct Pro	gca Ala	ggt Gly	gct Ala	gtc Val 500	aag Lys	cct Pro	tcg Ser	aaa Lys	gcc Ala 505	tca Ser	gac Asp	ttg Leu	1897
gat Asp	gac Asp 510	tac Tyr	ctg Leu	ggc Gly	ttc Phe	ctc Leu 515	ccc Pro	act Thr	ttt Phe	ggg Gly	ccc Pro 520	tgc Cys	tac Tyr	atc Ile	aac Asn	1945
ctc Leu 525	tat Tyr	ggc Gly	agt Ser	ccc Pro	aga Arg 530	gag Glu	ttc Phe	aca Thr	ggc Gly	ttc Phe 535	cca Pro	gac Asp	ccc Pro	tac Tyr	aca Thr 540	1993
gag Glu	ctc Leu	aac Asn	aca Thr	ggc Gly 545	aag Lys	ggg Gly	gaa Glu	ggt Gly	gtg Val 550	gct Ala	tat Tyr	cgt Arg	ggc Gly	cgg Arg 555	ctt Leu	2041
ctg Leu	ctc Leu	tcc Ser	ctg Leu 560	gag Glu	acc Thr	aag Lys	ctg Leu	gtg Val 565	gag Glu	cac His	agt Ser	gaa Glu	cag Gln 570	aag Lys	gtg Val	2089
gag Glu	gac Asp	ctt Leu 575	Pro	gcg Ala	gat Asp	gac Asp	atc Ile 580	Leu	cgg Arg	gtg Val	gag Glu	aag Lys 585	Tyr	ctt Leu	agg Arg	2137
agg Arg	cgc Arg 590	Lys	tac Tyr	tcc Ser	ctg Leu	ttt Phe 595	gcg Ala	gcc Ala	ttc Phe	tac Tyr	tca Ser 600	Ala	acc Thr	atg Met	ctg Leu	218
cag Gln 605	Asp	gtg Val	gat Asp	gat Asp	gcc Ala 610	Ile	cag Gln	ttt Phe	gag Glu	gto Val 615	Ser	atc Ile	ggg	aac Asn	tac Tyr 620	223
G1y	aac Asn	aag Lys	ttc Phe	gac Asp 625	Met	acc Thr	tgc Cys	ctg Leu	Pro 630	Leu	gcc Ala	tcc Ser	acc Thr	Thr 635	cag	228:
tac Tyr	ago Ser	cgt Arg	gca Ala 640	Val	ttt Phe	gac Asp	ggg Gly	tgc Cys 645	His	tac Tyr	tac Tyr	tac Tyr	cta Leu 650	Pro	tgg Trp	232

ggt. Gly	aac Asn	gtg Val 655	aaa Lys	cct Pro	gtg Val	gtg Val	gtg Val 660	ctg Leu	tca Ser	tcc Ser	tac Tyr	tgg Trp 665	gag Glu	gac Asp	atc Ile	2377
agc Ser	cat His 670	aga Arg	atc Ile	gag Glu	act Thr	cag Gln 675	aac Asn	cag Gln	ctg Leu	ctt Leu	680 Gly ggg	att Ile	gct Ala	gac Asp	cgg Arg	2425
ctg Leu 685	gaa Glu	gct Ala	ggc Gly	ctg Leu	gag Glu 690	cag Gln	gtc Val	cac His	ctg Leu	gcc Ala 695	ctg Leu	aag Lys	gcg Ala	cag Gln	tgc Cys 700	2473
tcc Ser	acg Thr	gag Glu	gac Asp	gtg Val 705	gac Asp	tcg Ser	ctg Leu	gtg Val	gct Ala 710	cag Gln	ctg Leu	acg Thr	gat Asp	gag Glu 715	ctc Leu	2521
atc Ile	gca Ala	ggc Gly	tgc Cys 720	agc Ser	cag Gln	cct Pro	ctg Leu	ggt Gly 725	gac Asp	atc Ile	cat His	gag Glu	aca Thr 730	ccc Pro	tct Ser	2569
gcc Ala	acc Thr	cac His 735	ctg Leu	gac Asp	cag Gln	tac Tyr	ctg Leu 740	tac Tyr	cag Gln	ctg Leu	cgc Arg	acc Thr 745	cat His	cac His	ctg Leu	2617
agc Ser	caa Gln 750	atc Ile	act Thr	gag Glu	gct Ala	gcc Ala 755	ctg Leu	gcc Ala	ctg Leu	aag Lys	ctc Leu 760	ggc Gly	cac His	agt Ser	gag Glu	2665
ctc Leu 765	cct Pro	gca Ala	gct Ala	ctg Leu	gag Glu 770	cag Gln	gcg Ala	gag Glu	gac Asp	tgg Trp 775	ctc Leu	ctg Leu	cgt Arg	ctg Leu	cgt Arg 780	2713
gcc Ala	ctg Leu	gca Ala	gag Glu	gag Glu 785	ccc Pro	cag Gln	aac Asn	agc Ser	ctg Leu 790	ccg Pro	gac Asp	atc Ile	gtc Val	atc Ile 795	tgg Trp	2761
atg Met	ctg Leu	cag Gln	gga Gly 800	gac Asp	aag Lys	cgt Arg	gtg Val	gca Ala 805	tac Tyr	cag Gln	cgg Arg	gtg Val	ccc Pro 810	gcc Ala	cac His	2809
caa Gln	gtc Val	ctc Leu 815	Phe	tcc Ser	cgg Arg	cgg Arg	ggt Gly 820	Ald	aac Asn	tac Tyr	tgt Cys	ggc Gly 825	aag Lys	aat Asn	tgt Cys	2857
Gly	aag Lys 830	Leu	cag Gln	aca Thr	atc	ttt Phe 835	ctg Leu	aaa Lys	tat Tyr	Pro	atg Met 840	GIU	aag Lys	gtg Val	cct Pro	2905
ggc Gly 845	Ala	cgg Arg	atg Met	cca Pro	gtg Val 850	GIn	ata Ile	cgg Arg	gto Val	aag Lys 855	Leu	tgg	ttt Phe	ggg Gly	ctc Leu 860	2953
tct Ser	gtg Val	gat Asp	gag Glu	aag Lys 865	Glu	tto Phe	aac Asn	cag Gln	ttt Phe 870	HIG	gag Glu	ggg Gly	aag Lys	ctg Leu 875	tct Ser	3001
gto Val	ttt Phe	gct Ala	gaa Glu 880	Thr	tat Tyr	gag Glu	aac Asr	gag Glu 885	Thr	aag Lys	tto Lev	gcc Ala	ctt Leu 890	• • • •	ggg	3049
aac Asi	tgg Tr	g ggg 6 Gly 899	Thi	acç Thr	g ggc	cto Lev	acc Thi	с туг	e ccc	aag Lys	ttt Phe	tct Ser 905	. voF	gtc Val	acg Thr	3097
gg¢ Gly	c aaq y Lys 910	3 Ile	aaq e Lys	g cta s Lev	a cco	aaq Lys 915	a Asi	e ago o Sei	tto Phe	e ego e Arç	920	) Sei	g gcc : Ala	ggc Gly	tgg Trp	3145

acc Thr 925	tgg Trp	gct Ala	gga Gly	Asp	tgg Trp 930	ttc Phe	gtg Val	tgt Cys	ccg Pro	gag Glu 935	aag Lys	act Thr	ctg Leu	ctc Leu	cat His 940	3193
gac Asp	atg Met	gac Asp	gcc Ala	ggt Gly 945	cac His	ctg Leu	agc Ser	ttc Phe	gtg Val 950	gaa Glu	gag Glu	gtg Val	ttt Phe	gag Glu 955	aac Asn	3241
cag Gln	acc Thr	cgg Arg	ctt Leu 960	ccc Pro	gga Gly	ggc Gly	cag Gln	tgg Trp 965	atc Ile	tac Tyr	atg Met	agt Ser	gac Asp 970	aac Asn	tac Tyr	3289
acc Thr	gat Asp	gtg Val 975	aac Asn	Gly ggg	gag Glu	aag Lys	gtg Val 980	ctt Leu	ccc Pro	aag Lys	gat Asp	gac Asp 985	att Ile	gag Glu	tgc Cys	3337
cca Pro	ctg Leu 990	Gly	tgg Trp	aag Lys	tgg Trp	gaa Glu 995	Asp	gag Glu	gaa Glu	tgg Trp	tcc Ser 1000	Thr	gac Asp	ctc Leu	aac Asn	3385
cgg Arg 1005	Ala	gtc Val	gat Asp	gag Glu	caa Gln 1010	Gly	tgg Trp	gag Glu	tat Tyr	agc Ser 1015	Ile	acc Thr	atc Ile	ccc Pro	ccg Pro 1020	3433
gag Glu	cgg Arg	aag Lys	ccg Pro	aag Lys 1025	His	tgg Trp	gtc Val	cct Pro	gct Ala 1030	gag Glu )	aag Lys	atg Met	tac Tyr	tac Tyr 1035	Thr	3481
cac His	cga Arg	cgg Arg	cgg Arg 1040	Arg	tgg Trp	gtg Val	cgc Arg	ctg Leu 1045	Arg	agg Arg	agg Arg	gat Asp	ctc Leu 1050	Ser	caa Gln	3529
atg Met	gaa Glu	gca Ala 1055	Leu	aaa Lys	agg Arg	cac His	agg Arg 1060	Gln	gcg Ala	gag Glu	gcg Ala	gag Glu 106	Gly	gag Glu	ggc Gly	3577
tgg Trp	gag Glu 1070	Tyr	gcc Ala	tct Ser	ctt Leu	ttt Phe 107	Gly	tgg Trp	aag Lys	ttc Phe	cac His 1080	Leu	gag Glu	tac Tyr	cgc Arg	3625
aag Lys 108	Thr	gat Asp	gcc Ala	ttc Phe	cgc Arg 1090	Arg	cgc Arg	ege Arg	tgg Trp	cgc Arg 109	Arg	cgc Arg	atg Met	gag Glu	cca Pro 1100	3673
ctg Leu	gag Glu	aag Lys	Thr	ggg Gly 110	Pro	gca Ala	gct Ala	gtg Val	ttt Phe 111	gcc Ala O	ctt Leu	gag Glu	ggg Gly	gcc Ala 111	Leu	3721
ggc Gly	ggc Gly	gtg Val	atg Met 112	Asp	gac Asp	aag Lys	agt Ser	gaa Glu 112	Asp	tcc Ser	atg Met	tcc Ser	gtc Val 113	Ser	acc Thr	3769
ttg Leu	agc Ser	ttc Phe 113	Gly	gtg Val	aac Asn	aga Arg	ccc Pro 114	Thr	att Ile	tcc Ser	tgc Cys	ata Ile 114	Phe	gac Asp	tat Tyr	3817
ggg Gly	aac Asn 115	Arg	tac Tyr	cat His	cta Leu	cgc Arg 115	Cys	tac Tyr	atg Met	tac Tyr	cag Gln 116	Ala	cgg Arg	gac Asp	ctg Leu	3865
gct																

ttc Phe	ctg Leu	cac His	Gln	agc Ser 1185	Gln	aag Lys	acg Thr	gtg Val	gtg Val 1190	Val	aag Lys	aac Asn	acc Thr	ctt Leu 1195	11011	3961
ccc Pro	acc Thr	tgg Trp	gac Asp 1200	Gln	acg Thr	ctc Leu	atc Ile	ttc Phe 1205	TYL	gag Glu	atc Ile	gag Glu	atc Ile 1210	1 110	ggc Gly	4009
Glu	Pro	Ala 1215	Thr	Val	Ala	Glu	1220	Pro )	PFO	Ser	116	gtg Val 1225	,	014	204	4057
Tyr	Asp 1230	His )	Asp	Thr	Tyr	1235	Ala	Asp	GIU	FILE	1240		Arg	0,5		4105
Cys 1245	Gln	Pro	Ser	Leu	1250	Arg	Mec	PIO	ALG	1255	5	tgg Trp			1260	4153
Thr	Arg	Gly	Ser	GIn 126	Pro	ser	GIĀ	Giu	1270	)	AIG	tct Ser	1	1279	5	4201
atc Ile	cag Gln	aga Arg	gag Glu 1280	Lys	ccg Pro	gcc Ala	atc Ile	cac His 128	UTO	att Ile	cct Pro	ggt Gly	ttt Phe 1290		gtg Val	4249
Gln	Glu	Thr 129	Ser 5	Arg	Ile	Leu	130	GIU O	ser	GIU	nsp	aca Thr 130	5	Deu	110	4297
Tyr	Pro 131	Pro 0	Pro	Gin	Arg	131	<b>А</b> Та	ASII	116	- Y -	132					4345
atc Ile 132	Lys	cca Pro	gcg Ala	ctc Leu	cag Gln 133	Arg	acc Thr	gcc Ala	atc Ile	gag Glu 133	116	ctg Leu	gca Ala	tgg Trp	ggc Gly 1340	4393
ctg Leu	cgg Arg	aac Asn	atg Met	aag Lys 134	Ser	tac Tyr	cag Gln	ctg Leu	gcc Ala 135	Wall	atc Ile	tcc Ser	tcc Ser	ccc Pro 135	001	4441
Leu	Val	Val	Glu 136	O Cys	GIĀ	GIŸ	GIN	136	5	GIII	Ser	Cys	137	0	agg Arg	4489
aac Asn	cto Leu	cgg Arg 137	Lys	aac Asn	ccc Pro	aac Asn	ttt Phe 138	Asp	atc Ile	tgc Cys	acc Thr	ctc Leu 138	File	atg Met	gaa Glu	4537
gtg Val	ato Met	Leu	ccc Pro	agg Arg	gag Glu	gag Glu 139	Ten	tac Tyr	tgc Cys	Pro	Pro 140	116	acc Thr	gtc Val	aag Lys	4585
gto Val 140	. Ile	gat Asp	aac Asn	cgc Arg	cag Gln 141	Phe	ggc Gly	cgc Arg	cgg Arg	cct Pro 141	, Agr	gtg Val	ggc	cag Gln	tgt Cys 1420	4633
acc Thi	ato	c cgc e Arç	tco Ser	cto Lev 142	GIU	ago Ser	tto Phe	c cto e Lev	tgt Cys 143	, voř	e ccc	tac Tyr	tco Ser	gcg Ala 143	g gag 1 Glu 35	4681
agt Sei	c cca	a tco Sei	e cca Pro	Glr	g ggt n Gly	ggc Gly	cca Pro	a gad Asp 144	o wal	gtç Val	g ago L Ser	cta Lev	t cto Leu 145		cct Pro	4729

ggg Gly	gaa Glu	gac Asp 1455	Val	ctc Leu	atc Ile	gac Asp	att Ile 1460	Asp	gac Asp	aag Lys	gag Glu	ccc Pro 1465	Leu	atc Ile	ccc Pro	4777
atc Ile	cag Gln 1470	Glu	gaa Glu	gag Glu	ttc Phe	atc Ile 1475	Asp	tgg Trp	tgg Trp	agc Ser	aaa Lys 1480	Phe	ttt Phe	gcc Ala	tcc Ser	4825
ata Ile 1489	ggg Gly	gag Glu	agg Arg	gaa Glu	aag Lys 1490	Cys	ggc Gly	tcc Ser	tac Tyr	ctg Leu 1495	Glu	aag Lys	gat Asp	ttt Phe	gac Asp 1500	4873
acc Thr	ctg Leu	aag Lys	gtc Val	tat Tyr 1505	Asp	aca Thr	cag Gln	ctg Leu	gag Glu 1510	Asn	gtg Val	gag Glu	gcc Ala	ttt Phe 1515	Glu	4921
ggc Gly	ctg Leu	tct Ser	gac Asp 1520	Phe	Cys tgt	aac Asn	acc Thr	ttc Phe 1525	Lys	ctg Leu	tac Tyr	cgg Arg	ggc Gly 1530	Lys	acg Thr	4969
cag Gln	gag Glu	gag Glu 1535	Thr	gaa Glu	gat Asp	cca Pro	tct Ser 1540	Val	att Ile	ggt Gly	gaa Glu	ttt Phe 1545	Lys	ggc Gly	ctc Leu	5017
ttc Phe	aaa Lys 1550	Ile	tat Tyr	ccc Pro	ctc Leu	cca Pro 1555	Glu	gac Asp	cca Pro	gcc Ala	atc Ile 1560	Pro	atg Met	ccc Pro	cca Pro	5065
aga Arg 156	cag Gln 5	ttc Phe	cac His	cag Gln	ctg Leu 1570	Ala	gcc Ala	cag Gln	gga Gly	ccc Pro 157	Gln	gag Glu	tgc Cys	ttg Leu	gtc Val 1580	5113
cgt Arg	atc Ile	tac Tyr	att Ile	gtc Val 158	Arg	gca Ala	ttt Phe	ggc Gly	ctg Leu 1590	Gln	ccc Pro	aag Lys	gac Asp	ccc Pro 159	Asn	5161
gga Gly	aag Lys	tgt Cys	gat Asp 160	Pro	tac Tyr	atc Ile	aag Lys	atc Ile 160	Ser	ata Ile	ggg Gly	aag Lys	aaa Lys 161	Ser	gtg Val	5209
agt Ser	gac Asp	cag Gln 161	Asp	aac Asn	tac Tyr	atc Ile	ccc Pro 162	Cys	acg Thr	ctg Leu	gag Glu	ccc Pro 162	Val	ttt Phe	gga Gly	5257
aag Lys	atg Met 163	ttc Phe 0	Glu	Leu	Thr	Cys	Thr	Leu	Pro	Leu	gag Glu 164	Lys	gac Asp	cta Leu	aag Lys	5305
ato Ile 164	act Thr 5	ctc Leu	tat Tyr	gac Asp	tat Tyr 165	Asp	ctc Leu	ctc Leu	tcc Ser	aag Lys 165	Asp	gaa Glu	aag Lys	atc Ile	ggt Gly 1660	5353
gag Glu	acg Thr	gtc Val	gtc Val	gac Asp 166	Leu	gag Glu	aac Asn	agg Arg	ctg Leu 167	Leu	tcc Ser	aag Lys	ttt Phe	999 Gly 167	Ala	5401
cgc Arg	tgt Cys	gga Gly	ctc Leu 168	Pro	cag Gln	acc Thr	tac Tyr	tgt Cys 168	Val	tct Ser	gga Gly	ccg Pro	aac Asn 169	Gln	tgg Trp	5449
cgg <b>Ar</b> g	gac Asp	cag Gln 169	Leu	cgc	Pro	tcc Ser	cag Gln 170	Leu	ctc Leu	cac His	ctc Leu	ttc Phe 170	Cys	cag Gln	cag Gln	5497

	Arg 1710	Val	Lys	Ala	Pro	vai 1715	Tyr	Arg	IIII	vaħ	1720	)	Mec		<b>01</b>	5545
gat Asp 1725	Lys	gaa Glu	tat Tyr	tcc Ser	att Ile 1730	Glu	gag Glu	ata Ile	gag Glu	gct Ala 1735	GIY	agg Arg	atc Ile	cca Pro	aac Asn 1740	5593
cca Pro	cac His	ctg Leu	ggc Gly	cca Pro 1745	vaı	gag Glu	gag Glu	cgt Arg	ctg Leu 1750	VIG	ctg Leu	cat His	gtg Val	ctt Leu 1755	<b></b>	5641
cag Gln	cag Gln	ggc Gly	ctg Leu 1760	Val	ccg Pro	gag Glu	cac His	gtg Val 1765	GIU	tca Ser	cgg Arg	ccc Pro	ctc Leu 1770	-1-	agc Ser	5689
ccc Pro	ctg Leu	cag Gln 1775	Pro	gac Asp	atc Ile	gag Glu	cag Gln 1780	GIY	aag Lys	ctg Leu	cag Gln	atg Met 1785		gtc Val	gac Asp	5737
cta Leu	ttt Phe 1790	Pro	aag Lys	gcc Ala	ctg Leu	ggg Gly 1795	Arg	cct Pro	gga Gly	cct Pro	ccc Pro 1800	ttc Phe	aac Asn	atc Ile	acc Thr	5785
cca Pro 180	Arg	aga Arg	gcc Ala	aga Arg	agg Arg 1810	Pne	ttc Phe	ctg Leu	cgt Arg	tgt Cys 181		atc Ile	tgg Trp	aat Asn	acc Thr 1820	5833
aga Arg	gat Asp	gtg Val	atc Ile	ctg Leu 182	Asp	gac Asp	ctg Leu	agc Ser	ctc Leu 183	IIII	ggg Gly	gag Glu	aag Lys	atg Met 183	Der	5881
gac Asp	att Ile	tat Tyr	gtg Val 184	Lys	ggt Gly	tgg Trp	atg Met	att Ile 184	GIY	ttt Phe	gaa Glu	gaa Glu	cac His 185	ت لات	caa Gln	5929
aag Lys	aca Thr	gac Asp 185	Val	cat His	tat Tyr	cgt Arg	tcc Ser 186	Leu	gga Gly	ggt Gly	gaa Glu	ggc Gly 186	ASII	ttc Phe	aac Asn	5977
tgg Trp	agg Arg 187	Phe	att Ile	ttc Phe	ccc Pro	ttc Phe 187	Asp	tac Tyr	ctg Leu	cca Pro	gct Ala 188	gag Glu O	caa Gln	gtc Val	tgt Cys	6025
acc Thr 188	Ile	gcc Ala	aag Lys	T.VS	aga	АІА	Pne	rrp	Arg	Leu	vah	aag Lys	T 11T	gag Glu	agc Ser 1900	6073
		cca Pro	gca Ala	cga Arg	var	gtg Val	ttc Phe	cag Gln	atc Ile 191		gac Asp	aat Asn	gac Asp	aag Lys 191	ttc Phe 5	6121
tcc Ser	ttt Phe	gat Asp	gat Asp 192	Phe	ctg Leu	ggc	tcc Ser	ctg Leu 192	GIII	cto Leu	gat Asp	ctc Leu	aac Asn 193	nry	atg Met	6169
ccc	aag Lys	cca Pro	) Ala	aag Lys	aca Thr	gcc Ala	aag Lys 194	rys	tgc Cys	tcc Ser	tto Lev	gac Asp 194	GIL	ctg Leu	gat Asp	6217
gat Asp	gct Ala	. Phe	cac His	cca Pro	gaa Glu	tgg Trp 195	Pne	gto Val	tco Ser	ctt Leu	ttt Phe 196	GIU	cag Gln	aaa Lys	aca Thr	6265
gto Val 196	L Lys	g gg¢	tgg Tr	g tgg Tr	9 ccc 9 Pro 197	Cys	gta Val	a gca L Ala	a gaa a Glu	gag Glu 197	T GT	gag Glu	aag Lys	aaa Lys	ata 3 Ile 1980	6313

ctg Leu	gcg Ala	ggc Gly	aag Lys	ctg Leu 1985	Glu	atg Met	acc Thr	ttg Leu	gag Glu 1990	11e	gta Val	gca Ala	gag Glu	agt Ser 1995	GIU	6361
cat His	gag Glu	gag Glu	cgg Arg 2000	Pro	gct Ala	ggc Gly	cag Gln	ggc Gly 2009	Arg	gat Asp	gag Glu	ccc Pro	aac Asn 2010	atg Met )	aac Asn	6409
cct Pro	aag Lys	ctt Leu 201	Glu	gac Asp	cca Pro	agg Arg	cgc Arg 2020	Pro	gac Asp	acc Thr	tcc Ser	ttc Phe 202	Leu	tgg Trp	ttt Phe	6457
acc Thr	tcc Ser 2030	Pro	tac Tyr	aag Lys	acc Thr	atg Met 203	Lys	ttc Phe	atc Ile	ctg Leu	tgg Trp 2040	Arg	cgt Arg	ttc Phe	cgg Arg	6505
tgg Trp 204	Āla	atc Ile	atc Ile	ctc Leu	ttc Phe 2050	Ile	atc Ile	ctc Leu	ttc Phe	atc Ile 205	Leu	ctg Leu	ctg Leu	ttc Phe	ctg Leu 2060	6553
gcc Ala	atc Ile	ttc Phe	atc Ile	tac Tyr 206	Ala	ttc Phe	ccg Pro	aac Asn	tat Tyr 2070	Ala	gcc Ala	atg Met	aag Lys	ctg Leu 207	Val	6601
	ccc Pro				ggac	tct (	cctg	ccct	gt a	gaag	gggc	c gt	gggg.	tece		6653
agge gtte acge	cctg gcta cttt	att aca ttt	gtcc tgga ggat	tgcc.	ag g	gtgg: gatc	gcag. accc	a cad	gaca ttcc	gatg atca	gac	cggc	cca ctc	CCCC	cctcct cccaga aaccca aaaaaa	6713 6773 6833 6893 6911

<210> 2 <211> 2080

<212> PRT

<213> Homo sapiens

<400> 2 Met Leu Arg Val Phe Ile Leu Tyr Ala Glu Asn Val His Thr Pro Asp 10 Thr Asp Ile Ser Asp Ala Tyr Cys Ser Ala Val Phe Ala Gly Val Lys 30 20 25 Lys Arg Thr Lys Val Ile Lys Asn Ser Val Asn Pro Val Trp Asn Glu 45 40 35 Gly Phe Glu Trp Asp Leu Lys Gly Ile Pro Leu Asp Gln Gly Ser Glu 55 Leu His Val Val Lys Asp His Glu Thr Met Gly Arg Asn Arg Phe 70 Leu Gly Glu Ala Lys Val Pro Leu Arg Glu Val Leu Ala Thr Pro Ser 90 85 Leu Ser Ala Ser Phe Asn Ala Pro Leu Leu Asp Thr Lys Lys Gln Pro 110 105 100 Thr Gly Ala Ser Leu Val Leu Gln Val Ser Tyr Thr Pro Leu Pro Gly 120 125 115 Ala Val Pro Leu Phe Pro Pro Pro Thr Pro Leu Glu Pro Ser Pro Thr 135 140 Leu Pro Asp Leu Asp Val Val Ala Asp Thr Gly Gly Glu Glu Asp Thr 155 150 Glu Asp Gln Gly Leu Thr Gly Asp Glu Ala Glu Pro Phe Leu Asp Gln 170 175 165 Ser Gly Gly Pro Gly Ala Pro Thr Thr Pro Arg Lys Leu Pro Ser Arg 190 185 Pro Pro Pro His Tyr Pro Gly Ile Lys Arg Lys Arg Ser Ala Pro Thr 200 195

Ser		Lys	Leu	Leu	Ser	Asp 215	Lys	Pro	Gln	Asp	Phe 220	Gln	Ile	Arg	Val
	Val	Ile	Glu	Gly	Arg 230	Gln	Leu	Pro	Gly	Val 235	Asn	Ile	Lys	Pro	Val 240
	_			245	Ala				250	Arg				233	
_			260	Leu				265		Phe			2/0		
		275	Glu				280			Phe		200			
	200	Ser				295				Gly	300				
205					310					Ala 315					J 2 U
				マクち					330	Ala				333	
			310					345		Gly			220		
		255					360			Asp		303			
	270					375				Ala	300				
205					าจก					Asp 395					400
				405					410	Asn				417	
_			420					425		Lys			430		
		4 2 E					440			Asn		443			
	4EO					455				Met	400				
400					17N					Val 475 Glu					400
				125					490	Asp				499	
			ここ こうしゅう こうしゅう こうしゅう こうしゅう こうしゅう こうしゅう こうしゅう こうしゅう こうしゅう しゅう しゅうしゅう しゅう					วบว		Ile			210		
		E15					520			Tyr		323			
	E 2 A					535	,			Arg	240				
F 4 E					ちちい					555 Lys					500
				C C C					5 / U	,				3/3	
			E 9 /	i .				363		Leu					
		605	:				600	,				000			Asp Phe
	610	١				615	5				020				Ala
625	:				630					ರವಾ					040
				645					שטע	,				000	
			661	)				665	)				6,0		Ile Glv
		671	Ε.				680	)				003	,		Gly
	600	`				695	5				/00	,			Asp
700					710	)				/15	)				Cys 720 Leu
Sei	Gl:	n Pro	o Le	3 Gly 725	Asī	) TT6	; n18	י פדו	730	)	, ser	C		735	Leu

Asp Gln Tyr Leu Tyr Gln Leu Arg Thr His His Leu Ser Gln Ile Thr Glu Ala Ala Leu Ala Leu Lys Leu Gly His Ser Glu Leu Pro Ala Ala Leu Glu Gln Ala Glu Asp Trp Leu Leu Arg Leu Arg Ala Leu Ala Glu Glu Pro Gln Asn Ser Leu Pro Asp Ile Val Ile Trp Met Leu Gln Gly Asp Lys Arg Val Ala Tyr Gln Arg Val Pro Ala His Gln Val Leu Phe Ser Arg Arg Gly Ala Asn Tyr Cys Gly Lys Asn Cys Gly Lys Leu Gln Thr Ile Phe Leu Lys Tyr Pro Met Glu Lys Val Pro Gly Ala Arg Met Pro Val Gln Ile Arg Val Lys Leu Trp Phe Gly Leu Ser Val Asp Glu Lys Glu Phe Asn Gln Phe Ala Glu Gly Lys Leu Ser Val Phe Ala Glu Thr Tyr Glu Asn Glu Thr Lys Leu Ala Leu Val Gly Asn Trp Gly Thr Thr Gly Leu Thr Tyr Pro Lys Phe Ser Asp Val Thr Gly Lys Ile Lys Leu Pro Lys Asp Ser Phe Arg Pro Ser Ala Gly Trp Thr Trp Ala Gly Asp Trp Phe Val Cys Pro Glu Lys Thr Leu Leu His Asp Met Asp Ala Gly His Leu Ser Phe Val Glu Glu Val Phe Glu Asn Gln Thr Arg Leu Pro Gly Gly Gln Trp Ile Tyr Met Ser Asp Asn Tyr Thr Asp Val Asn Gly Glu Lys Val Leu Pro Lys Asp Asp Ile Glu Cys Pro Leu Gly Trp Lys Trp Glu Asp Glu Glu Trp Ser Thr Asp Leu Asn Arg Ala Val Asp Glu Gln Gly Trp Glu Tyr Ser Ile Thr Ile Pro Pro Glu Arg Lys Pro Lys His Trp Val Pro Ala Glu Lys Met Tyr Tyr Thr His Arg Arg 1025 1030 1035 1040 Arg Trp Val Arg Leu Arg Arg Arg Asp Leu Ser Gln Met Glu Ala Leu Lys Arg His Arg Gln Ala Glu Ala Glu Gly Glu Gly Trp Glu Tyr Ala 1060 1065 1070 Ser Leu Phe Gly Trp Lys Phe His Leu Glu Tyr Arg Lys Thr Asp Ala Phe Arg Arg Arg Trp Arg Arg Met Glu Pro Leu Glu Lys Thr Gly Pro Ala Ala Val Phe Ala Leu Glu Gly Ala Leu Gly Gly Val Met Asp Asp Lys Ser Glu Asp Ser Met Ser Val Ser Thr Leu Ser Phe Gly Val Asn Arg Pro Thr Ile Ser Cys Ile Phe Asp Tyr Gly Asn Arg Tyr 1140 1145 1150 His Leu Arg Cys Tyr Met Tyr Gln Ala Arg Asp Leu Ala Ala Met Asp Lys Asp Ser Phe Ser Asp Pro Tyr Ala Ile Val Ser Phe Leu His Gln Ser Gln Lys Thr Val Val Val Lys Asn Thr Leu Asn Pro Thr Trp Asp Gln Thr Leu Ile Phe Tyr Glu Ile Glu Ile Phe Gly Glu Pro Ala Thr Val Ala Glu Gln Pro Pro Ser Ile Val Val Glu Leu Tyr Asp His Asp Thr Tyr Gly Ala Asp Glu Phe Met Gly Arg Cys Ile Cys Gln Pro Ser 1235 1240 Leu Glu Arg Met Pro Arg Leu Ala Trp Phe Pro Leu Thr Arg Gly Ser 

Gln Pro Ser Gly Glu Leu Leu Ala Ser Phe Glu Leu Ile Gln Arg Glu 1265 1270 1275 1280 Lys Pro Ala Ile His His Ile Pro Gly Phe Glu Val Gln Glu Thr Ser 1290 1285 Arg Ile Leu Asp Glu Ser Glu Asp Thr Asp Leu Pro Tyr Pro Pro Pro 1300 1305 1310 Gln Arg Glu Ala Asn Ile Tyr Met Val Pro Gln Asn Ile Lys Pro Ala 1325 1320 1315 Leu Gln Arg Thr Ala Ile Glu Ile Leu Ala Trp Gly Leu Arg Asn Met 1335 Lys Ser Tyr Gln Leu Ala Asn Ile Ser Ser Pro Ser Leu Val Val Glu 1355 1350 Cys Gly Gly Gln Thr Val Gln Ser Cys Val Ile Arg Asn Leu Arg Lys 1375 1370 1365 Asn Pro Asn Phe Asp Ile Cys Thr Leu Phe Met Glu Val Met Leu Pro 1380 1385 1390 Arg Glu Glu Leu Tyr Cys Pro Pro Ile Thr Val Lys Val Ile Asp Asn 1405 1395 1400 Arg Gln Phe Gly Arg Arg Pro Val Val Gly Gln Cys Thr Ile Arg Ser 1420 1415 Leu Glu Ser Phe Leu Cys Asp Pro Tyr Ser Ala Glu Ser Pro Ser Pro 1425 1430 1435 1446 Gln Gly Gly Pro Asp Asp Val Ser Leu Leu Ser Pro Gly Glu Asp Val 1445 1450 Leu Ile Asp Ile Asp Asp Lys Glu Pro Leu Ile Pro Ile Gln Glu Glu 1470 1465 1460 Glu Phe Ile Asp Trp Trp Ser Lys Phe Phe Ala Ser Ile Gly Glu Arg 1485 1475 1480 Glu Lys Cys Gly Ser Tyr Leu Glu Lys Asp Phe Asp Thr Leu Lys Val 1500 1495 1490 Tyr Asp Thr Gln Leu Glu Asn Val Glu Ala Phe Glu Gly Leu Ser Asp 1515 1510 Phe Cys Asn Thr Phe Lys Leu Tyr Arg Gly Lys Thr Gln Glu Glu Thr 1525 1530 1535 Glu Asp Pro Ser Val Ile Gly Glu Phe Lys Gly Leu Phe Lys Ile Tyr 1545 1550 1540 Pro Leu Pro Glu Asp Pro Ala Ile Pro Met Pro Pro Arg Gln Phe His 1565 1560 1555 Gln Leu Ala Ala Gln Gly Pro Gln Glu Cys Leu Val Arg Ile Tyr Ile 1570 1575 1580 Val Arg Ala Phe Gly Leu Gln Pro Lys Asp Pro Asn Gly Lys Cys Asp 1590 1595 Pro Tyr Ile Lys Ile Ser Ile Gly Lys Lys Ser Val Ser Asp Gln Asp 1605 1610 1615 Asn Tyr Ile Pro Cys Thr Leu Glu Pro Val Phe Gly Lys Met Phe Glu 1625 1620 Leu Thr Cys Thr Leu Pro Leu Glu Lys Asp Leu Lys Ile Thr Leu Tyr 1645 1640 1635 Asp Tyr Asp Leu Leu Ser Lys Asp Glu Lys Ile Gly Glu Thr Val Val 1655 1660 Asp Leu Glu Asn Arg Leu Leu Ser Lys Phe Gly Ala Arg Cys Gly Leu 1670 1675 Pro Gln Thr Tyr Cys Val Ser Gly Pro Asn Gln Trp Arg Asp Gln Leu 1690 1695 1685 Arg Pro Ser Gln Leu Leu His Leu Phe Cys Gln Gln His Arg Val Lys 1710 1705 1700 Ala Pro Val Tyr Arg Thr Asp Arg Val Met Phe Gln Asp Lys Glu Tyr 1725 1720 1715 Ser Ile Glu Glu Ile Glu Ala Gly Arg Ile Pro Asn Pro His Leu Gly 1740 1735 Pro Val Glu Glu Arg Leu Ala Leu His Val Leu Gln Gln Gln Gly Leu 1755 1750 Val Pro Glu His Val Glu Ser Arg Pro Leu Tyr Ser Pro Leu Gln Pro 1770 1775 1765 Asp Ile Glu Gln Gly Lys Leu Gln Met Trp Val Asp Leu Phe Pro Lys 1785

## 13/68

```
Ala Leu Gly Arg Pro Gly Pro Pro Phe Asn Ile Thr Pro Arg Arg Ala
                           1800
         1795
Arg Arg Phe Phe Leu Arg Cys Ile Ile Trp Asn Thr Arg Asp Val Ile
                                                1820
    1810
                          1815
Leu Asp Asp Leu Ser Leu Thr Gly Glu Lys Met Ser Asp Ile Tyr Val
                                            1835
                      1830
1825
Lys Gly Trp Met Ile Gly Phe Glu Glu His Lys Gln Lys Thr Asp Val
                                                             1855
                                       1850
                 1845
His Tyr Arg Ser Leu Gly Gly Glu Gly Asn Phe Asn Trp Arg Phe Ile
                                                         1870
                                   1865
             1860
Phe Pro Phe Asp Tyr Leu Pro Ala Glu Gln Val Cys Thr Ile Ala Lys
                              1880
                                                    1885
         1875
Lys Asp Ala Phe Trp Arg Leu Asp Lys Thr Glu Ser Lys Ile Pro Ala
                                                1900
                          1895
Arg Val Val Phe Gln Ile Trp Asp Asn Asp Lys Phe Ser Phe Asp Asp
                                            1915
                      1910
1905
Phe Leu Gly Ser Leu Gln Leu Asp Leu Asn Arg Met Pro Lys Pro Ala
                                       1930
                                                              1935
                  1925
Lys Thr Ala Lys Lys Cys Ser Leu Asp Gln Leu Asp Asp Ala Phe His
                                                         1950
                                   1945
             1940
Pro Glu Trp Phe Val Ser Leu Phe Glu Gln Lys Thr Val Lys Gly Trp
                                                     1965
                               1960
         1955
Trp Pro Cys Val Ala Glu Glu Gly Glu Lys Lys Ile Leu Ala Gly Lys
                                                1980
                          1975
    1970
Leu Glu Met Thr Leu Glu Ile Val Ala Glu Ser Glu His Glu Glu Arg
                                            1995
                      1990
1985
Pro Ala Gly Gln Gly Arg Asp Glu Pro Asn Met Asn Pro Lys Leu Glu
                                                             2015
                  2005
                                       2010
Asp Pro Arg Arg Pro Asp Thr Ser Phe Leu Trp Phe Thr Ser Pro Tyr
                                                         2030
                                   2025
              2020
Lys Thr Met Lys Phe Ile Leu Trp Arg Arg Phe Arg Trp Ala Ile Ile
                               2040
                                                     2045
         2035
Leu Phe Ile Ile Leu Phe Ile Leu Leu Phe Leu Ala Ile Phe Ile
                                                 2060
                           2055
Tyr Ala Phe Pro Asn Tyr Ala Ala Met Lys Leu Val Lys Pro Phe Ser
                                                                  2080
                      2070
                                            2075
2065
       <210> 3
       <211> 5915
       <212> DNA
       <213> Homo sapiens
       <400> 3
tcgaccgccc agccaggtgc aaaatgccgt gtcattggga gactccgcag ccggagcatt
agattacage tegacggage tegggaaggg eggegggggt ggaagatgag cagaageeee tgttetegga acgeeggetg acaagegggg tgagegeagg eggggegggg acceageeta
                                                                             120
                                                                             180
gcccactgga gcagccggg gtggcccgtt cccctttaag agcaactgct ctaagccagg
                                                                             240
agccagagat tegageegge etegeecage cageectete cagegagggg acceacaage
                                                                             300
ggcgcctcgg ccctcccgac ctttccgagc cctctttgcg ccctgggcgc acggggccct
                                                                             360
acacgegeca ageatgetga gggtetteat cetetatgee gagaaegtee acacaceega
                                                                             420
caccgacate agegatgeet actgeteege ggtgtttgea ggggtgaaga agagaaceaa
                                                                             480
agtcatcaag aacagcgtga accetgtatg gaatgaggga titgaatggg accetcaaggg
                                                                             540
catececety gaccaggget etgagettea tgtggtggte aaagaccatg agacgatggg
                                                                             600
gaggaacagg ttectggggg aagccaaggt cccactccga gaggtecteg ccaccctag tetgteegec agetteaatg ecceetget ggacaccaag aagcagecca caggggeete
                                                                             660
                                                                             720
getggteetg caggtgteet acacaceget geetggaget gtgeecetgt tecegeeece
                                                                             780
tactectetg gagecetece egactetgee tgacetggat gtagtggeag acacaggagg agaggaagac acagaggace agggacteae tggagatgag geggagecat teetggatea
                                                                             840
                                                                             900
```

aagcggagge cegggggete ceaceacec aaggaaacta cetteacgte etecgeecea ctacccggg atcaaagaa agcgaagtgc gcctacatct agaaagctgc tgtcagacaa

accgcaggat ttccagatca gggtccaggt gatcgagggg cgccagctgc cggggggtgaa

catcaageet gtggteaagg ttaccgetge agggeagace aageggaege ggatecacaa

gggaaacage ccactettea atgagactet tttetteaac ttgtttgaet eteetgggga

gctgtttgat gagcccatct ttatcacggt ggtagactct cgttctctca ggacagatgc

tetecteggg gagtteegga tggaegtggg caccatttae agagageece ggeaegeeta teteaggaag tggetgetge teteagaeee tgatgaette tetgetgggg ecagaggeta

cctgaaaaca agcctttgtg tgctggggcc tggggacgaa gcgcctctgg agagaaaaga

BNSDOCID: <WO\_\_\_\_0011157A1\_I\_>

ccctctgaa gacaaggagg acattgaaag caacctgctc cggcccacag gcgtagccct gcgaggagcc cacttctgcc tgaaggtctt ccgggccgag gacttgccgc agatggacga 1500 1560 tgccgtgatg gacaacgtga aacagatctt tggcttcgag agtaacaaga agaacttggt ggaccecttt gtggaggtca gctttgcggg gaaaatgctg tgcagcaaga tcttggagaa gacggccaac cctcagtgga accagaacat cacactgcct gccatgtttc cctccatgtg 1620 1680 1740 cgaaaaaatg aggattegta teatagaetg ggacegeetg acteacaatg acategtgge taccacetae etgagtatgt egaaaatete tgeeeetgga ggagaaatag aagaggagee 1800 1860 tgcaggtgct gtcaagcctt cgaaagcctc agacttggat gactacctgg gcttcctcc cacttttggg ccctgctaca tcaacctcta tggcagtccc agagagttca caggcttccc 1920 1980 agacccctac acagagetca acacaggeaa gggggaaggt gtggettate gtggeegget 2040 tetgetetee etggagacca agetggtgga geacagtgaa cagaaggtgg aggacettee 2100 tgcggatgac atcctccggg tggagaagta ccttaggagg cgcaagtact ccctgtttgc 2160 ggccttctac tcagccacca tgctgcagga tgtggatgat gccatccagt ttgaggtcag 2220 catcgggaac tacgggaaca agttcgacat gacctgcctg ccgctggcct ccaccactca gtacagccgt gcagtctttg acgggtgcca ctactactac ctacctggg gtaacgtgaa acctgtggtg gtgctgtcat cctactggga ggacatcagc catagaatcg agactcagaa 2280 2340 2400 2460 ccagctgctt gggattgctg accggctgga agctggcctg gagcaggtcc acctggcct gaaggegeag tgetecaegg aggaegtgga etegetggtg geteagetga eggatgaget 2520 categeagge tgeagecage etetgggtga catecatgag acacectetg ceacceacet 2580 ggaccagtac ctgtaccage tgcgcaccca tcacctgage caaatcactg aggetgeeet 2640 2700 ggccctgaag ctcggccaca gtgagctccc tgcagctctg gagcaggcgg aggactggct cetgegtetg egtgeeetgg cagaggagee ceagaacage etgeeggaca tegteatetg 2760 gatgctgcag ggagacaagc gtgtggcata ccagcgggtg cccgcccacc aagtcctctt ctcccggcgg ggtgccaact actgtggcaa gaattgtggg aagctacaga caatctttct 2820 2880 gaaatatccg atggagaagg tgcctggcgc ccggatgcca gtgcagatac gggtcaagct gtggtttggg ctctctgtgg atgagaagga gttcaaccag tttgctgagg ggaagctgtc 2940 3000 tgtctttgct gaaacctatg agaacgagac taagttggcc cttgttggga actggggcac 3060 aacgggcctc acctacccca agttttctga cgtcacgggc aagatcaagc tacccaagga 3120 cagetteege ceeteggeeg getggaectg ggetggagat tggttegtgt gteeggagaa gaetetgete catgacatgg acgeeggtea cetgagette gtggaagagg tgtttgagaa ceagaecegg etteeeggag geeagtggat etacatgagt gaeaactaca eegatgtgaa 3180 3240 3300 cggggagaag gtgcttccca aggatgacat tgagtgccca ctgggctgga agtgggaaga 3360 tgaggaatgg tccacagacc tcaaccgggc tgtcgatgag caaggctggg agtatagcat caccatccc ccggagcgga agccgaagca ctgggtccct gctgagaaga tgtactacac 3420 3480 acaccgacgg cggcgctggg tgcgcctgcg caggagggat ctcagccaaa tggaagcact gaaaaggcac aggcaggcgg aggcggaggg cgagggctgg gagtacgcct ctctttttgg 3540 3600 ctggaagtte cacetegagt accgcaagac agatgcette egeegeegee getggegeegtegeatggag ccaetggaga agacgggee tgcagetgtg tttgeeettg agggggeeet 3660 3720 gggcgcgtg atggatgaca agagtgaaga ttccatgtcc gtctccacct tgagcttcgg 3780 tgtgaacaga cccacgattt cctgcatatt cgactatggg aaccgctacc atctacgctg 3840 ctacatgtac caggeceggg acctggetge gatggacaag gactetttt etgateeta 3900 tgccatcgtc tccttcctgc accagagcca gaagacggtg gtggtgaaga acacccttaa ccccacctgg gaccagacgc tcatcttcta cgagatcgag atctttggcg agccggccac 3960 4020 agttgctgag caaccgccca gcattgtggt ggagctgtac gaccatgaca cttatggtgc 4080 agacgagttt atgggtcgct gcatctgtca accgagtctg gaacggatgc cacggctggc 4140 ctggttccca ctgacgaggg gcagccagcc gtcgggggag ctgctggcct cttttgagct 4200 catccagaga gagaagccgg ccatccacca tattcctggt tttgaggtgc aggagacatc 4260 aaggateetg gatgagtetg aggacacaga cetgeeetae ceaceacee agagggagge caacatetae atggtteete agaacateaa gecagegete cagegtaeeg ceategagat 4320 4380 cctggcatgg ggcctgcgga acatgaagag ttaccagctg gccaacatct cctcccccag cctcgtggta gagtgtgggg gccagacggt gcagtcctgt gtcatcagga acctccggaa 4440 4500 gaaccccaac tttgacatet gcaccctctt catggaagtg atgctgccca gggaggagct 4560 ctactgccc cccatcaccg tcaaggtcat cgataaccgc cagtttggcc gccggcctgt 4620 ggtgggccag tgtaccatcc gctccctgga gagcttcctg tgtgacccct actcggcgga 4680 gagtecatec ccacagggtg geccagacga tgtgagecta etcagtectg gggaagaegt 4740 getcategae attgatgaca aggageceet catececate caggaggaag agtteatega 4800 ttggtggagc aaattetttg cetecatagg ggagagggaa aagtgegget cetacetgga 4860 gaaggatttt gacaccttga aggtctatga cacacagctg gagaatgtgg aggcctttga gggcctgtct gacttttgta acaccttcaa gctgtaccgg ggcaagacgc aggaggagac 4920 4980 agaagatcca tctgtgattg gtgaatttaa gggcctcttc aaaatttatc ccctcccaga agacccagcc atccccatgc ccccaagaca gttccaccag ctggccgccc agggacccca ggagtgcttg gtccgtatct acattgtccg agcatttggc ctgcagccca aggaccccaa tcgaaagtgt gatccttaca tcaagatctc catagggaag aaatcagtga gtgaccagga 5040 5100 5160 5220 taactacatc ccctgcacgc tggagcccgt atttggaaag atgttcgagc tgacctgcac tctgcctctg gagaaggacc taaagatcac tctctatgac tatgacctcc tctccaagga cgaaaagatc ggtgagacgg tcgtcgacct ggagaacagg ctgctgtcca agtttggggc 5280 5340 5400 tegetgtgga eteccacaga ectactgtgt etetggaceg aaccagtgge gggaccaget 5460

```
ccgccctcc cagctcctcc acctcttctg ccagcagcat agagtcaagg cacctgtgta
                                                                                      5520
ccggacagac cgtgtaatgt ttcaggataa agaatattcc attgaagaga tagaggctgg
                                                                                      5580
caggatecca aacccacace tgggeccagt ggaggagegt etggetetge atgtgettea geageaggge etggtecegg ageaegtgga gteaeggece etetaeagee ecettgageeggaget tgeagatgt ggtegaceta ttteegaagg ecetgggegggeggetggaget ecetteaaca teaececag ggaggeaga aggttttee tgegttgtat
                                                                                      5640
                                                                                      5700
                                                                                      5760
                                                                                      5820
tatctggaat accagagatg tgatcctgga tgacctgagc ctcacggggg agaagatgag
                                                                                      5880
                                                                                      5915
cgacatttat gtgaaaggtt ggatgattgg ctttg
        <210> 4
        <211> 20
         <212> DNA
         <213> Homo sapiens
         <400> 4
                                                                                          20
 tgggacctca aagggcatcc
         <210> 5
         <211> 20
<212> DNA
         <213> Homo sapiens
         <400> 5
                                                                                          20
 accatgctgt aggatgtgga
         <210> 6
         <211> 20
         <212> DNA
         <213> Homo sapiens
         <400> 6
                                                                                          20
 gggaggtgaa gcaacttcaa
         <210> 7
         <211> 20
         <212> DNA
         <213> Homo sapiens
         <400> 7
                                                                                          20
  ctcacggggt agaagatgag
         <210> 8
         <211> 20
         <212> DNA
         <213> Homo sapiens
         <400> 8
                                                                                          20
  cagggccgag atgagcccaa
         <210> 9
         <211> 20
         <212> DNA
         <213> Homo sapiens
         <400> 9
                                                                                          20
  acatcaaggg tcctggatga
         <210> 10
         <211> 20
          <212> DNA
          <213> Homo sapiens
          <400> 10
                                                                                          20
  ctgtggcggt gtttccggtg
          <210> 11
          <211> 20
```

```
<212> DNA
         <213> Homo sapiens
         <400> 11
                                                                                                  20
 acagacgtgc gttatcgttc
         <210> 12
         <211> 20
         <212> DNA
         <213> Homo sapiens
         <400> 12
                                                                                                  20
 aagactgagc aaaatcccag
         <210> 13
         <211> 6912
         <212> DNA
         <213> Homo sapiens
         <400> 13
tegacegece agecaggtge aaaatgeegt gteattggga gaeteegeag eeggageatt
                                                                                                 60
agattacage tegacggage tegggaaggg eggegggggt ggaagatgag cagaageece
                                                                                                120
180
gcccactgga gcagccgggg gtggcccgtt cccctttaag agcaactgct ctaagccagg agccagagat tcgagccgc ctcgccagc cagccctctc cagcgagggg acccacaagc ggcgcctcgg ccctcccgac ctttccgagc cctctttgcg ccctgggcgc acggggccct
                                                                                                240
                                                                                                300
                                                                                                360
acacgegeca ageatgetga gggtetteat cetetatgee gagaaegtee acacaceega
                                                                                                420
caccgacate agcgatgcct actgctccgc ggtgtttgca ggggtgaaga agagaaccga agtcatcaag aacagcgtga acctgtatg gaatgaggga tttgaatggg acctcaaagg gcatcccct ggaccagggc tctgagcttc atgtggtggt caaagaccat gagacgatgg
                                                                                                480
                                                                                                540
                                                                                                600
ggaggaacag gtteetgggg gaagecaagg teceaeteeg agaggteete gecaeeeeta gtetgteege cagetteaat geceeetge tggacaccaa gaageageee acaggggeet
                                                                                                660
                                                                                                720
egetggteet geaggtgtee tacacacege tgeetggage tgtgeecetg tteeegeece etacteetet ggageeetee eegactetge etgacetgga tgtagtggea gacacaggag
                                                                                                780
                                                                                                840
gagaggaaga cacagaggac cagggactca ctggagatga ggcggagcca ttcctggatc
                                                                                                900
aaagcggagg cccgggggct cccaccaccc caaggaaact accttcacgt cctccgcccc
                                                                                                960
actacccgg gatcaaaaga aagcgaagtg cgcctacatc tagaaagctg ctgtcagaca aaccgcagga tttccagatc agggtccagg tgatcgagg gcgccagctg ccggggggaacaagcgcgaacagggaaaaag cccactcttc aatgagactc tttcttcaa cttgtttgac tctcctgggg
                                                                                               1020
                                                                                               1080
                                                                                               1140
                                                                                              1200
agetgtttga tgageccate tttatcaegg tggtagaete tegttetete aggacagatg
                                                                                               1260
ctetectegg ggagtteegg atggaegtgg gcaccattta cagagagece eggeaegeet ateteaggaa gtggetgetg eteteagace etgatgaett etetgetggg gecagagget
                                                                                               1320
                                                                                               1380
acctgaaaac aagcetttgt gtgctgggge ctggggacga agcgeetetg gagagaaaag
                                                                                               1440
accectetga agacaaggag gacattgaaa geaacetget eeggeecaca ggegtageee tgegaggage ceaettetge etgaaggtet teegggeega ggacttgeeg eagatggaeg
                                                                                               1500
                                                                                               1560
atgccgtgat ggacaacgtg aaacagatct ttggcttcga gagtaacaag aagaacttgg
                                                                                               1620
tggacccctt tgtggaggtc agctttgcgg ggaaaatgct gtgcagcaag atcttggaga agacggccaa ccctcagtgg aaccagaaca tcacactgcc tgccatgttt ccctccatgt
                                                                                               1680
                                                                                               1740
gcgaaaaat gaggattegt atcatagact gggacegeet gaetcacaat gacategtgg
                                                                                               1800
 ctaccaccta cctgagtatg tcgaaaatct ctgcccctgg aggagaaata gaagaggagc
                                                                                               1860
 ctgcaggtgc tgtcaagcct tcgaaagcct cagacttgga tgactacctg ggcttcctcc ccacttttgg gccctgctac atcaacctct atggcagtcc cagagagttc acaggcttcc
                                                                                               1920
                                                                                               1980
 cagacccta cacagagete aacacaggea agggggaagg tgtggettat cgtggccgge
                                                                                               2040
 ttctgctctc cctggagacc aagctggtgg agcacagtga acagaaggtg gaggaccttc
                                                                                               2100
 ctgcggatga catcctccgg gtggagaagt accttaggag gcgcaagtac tccctgtttg
                                                                                               2160
 eggeetteta etcagecace atgetgtagg atgtggatga tgecatecag tttgaggtea
                                                                                               2220
 gcatcgggaa ctacgggaac aagttcgaca tgacctgcct gccgctggcc tccaccactc
                                                                                               2280
 agtacagecg tgcagtettt gaegggtgee actactaeta ectaecetgg ggtaacgtga
                                                                                               2340
 aacctgtggt ggtgctgtca tcctactggg aggacatcag ccatagaatc gagactcaga
                                                                                               2400
 accagetget tgggattget gaccggetgg aagetggeet ggageaggte cacetggeee
                                                                                               2460
 tgaaggegea gtgetecaeg gaggaegtgg actegetggt ggeteagetg acggatgage
                                                                                               2520
 tcatcgcagg ctgcagccag cctctgggtg acatccatga gacaccctct gccacccacc tggaccagta cctgtaccag ctgcgcaccc atcacctgag ccaaatcact gaggctgccc
                                                                                               2580
                                                                                               2640
 tggccctgaa gctcggccac agtgagctcc ctgcagctct ggagcaggcg gaggactggc tcctgcgtct gcgtgccctg gcagaggagc cccagaacag cctgccggac atcgtcatct
                                                                                               2700
                                                                                               2760
 ggatgctgca gggagacaag cgtgtggcat accagegggt geeegeecae caagteetet
                                                                                               2820
```

tctcccggcg gggtgccaac tactgtggca agaattgtgg gaagctacag acaatctttc tgaaatatcc gatggagaag gtgcctggcg cccggatgcc agtgcagata cgggtcaagc 2880 2940 3000 tgtggtttgg gctctctgtg gatgagaagg agttcaacca gtttgctgag gggaagctgt ctgtctttgc tgaaacctat gagaacgaga ctaagttggc ccttgttggg aactggggca 3060 3120 caacgggcct cacctacccc aagttttctg acgtcacggg caagatcaag ctacccaagg acagetteeg ececteggee ggetggaeet gggetggaga ttggttegtg tgteeggaga 3180 3240 agactotgot coatgacatg gacgooggto acctgagott cgtggaagag gtgtttgaga accagacceg gettecegga ggecagtgga tetacatgag tgacaactae accgatgtga acgggggagaa ggtgettece aaggatgaca ttgagtgece actgggetgg aagtgggaag 3300 3360 atgaggaatg gtccacagac ctcaaccggg ctgtcgatga gcaaggctgg gagtatagca tcaccatccc cccggagcgg aagccgaagc actgggtccc tgctgagaag atgtactaca 3420 3480 cacaccgacg gcggcgctgg gtgcgcctgc gcaggaggga tctcagccaa atggaagcac 3540 tgaaaaggca caggcaggcg gaggcggagg gcgagggctg ggagtacgcc tctcttttg gctggaagtt ccacctcgag taccgcaaga cagatgcctt ccgccgccgc cgctggcgcc 3600 3660 gtcgcatgga gccactggag aagacggggc ctgcagctgt gtttgccctt gagggggccc 3720 tgggcggcgt gatggatgac aagagtgaag attccatgtc cgtctccacc ttgagcttcg 3780 gtgtgaacag acccacgatt tcctgcatat tcgactatgg gaaccgctac catctacgct 3840 gctacatgta ccaggcccgg gacctggctg cgatggacaa ggactctttt tctgatccct atgccatcgt ctccttcctg caccagagcc agaagacggt ggtggtgaag aacaccctta 3900 3960 acccacctg ggaccagacg ctcatcttct acgagatcga gatctttggc gagccggcca 4020 cagttgctga gcaaccgccc agcattgtgg tggagctgta cgaccatgac acttatggtg 4080 cagacgagtt tatgggtcgc tgcatctgtc aaccgagtct ggaacggatg ccacggctgg cctggttccc actgacgagg ggcagccagc cgtcgggga gctgctggcc tcttttgagc tcatccagag agagaagccg gccatccacc atattcctgg ttttgaggtg caggagacat caaggatcct ggatgagtct gaggacacag acctgcccta cccaccacc cagagggagg 4140 4200 4260 4320 ccaacatcta catggttcct cagaacatca agccagcgct ccagcgtacc gccatcgaga 4380 tcctggcatg gggcctgcgg aacatgaaga gttaccagct ggccaacatc tcctcccca 4440 gcctcgtggt agagtgtggg ggccagacgg tgcagtcctg tgtcatcagg aacctccgga agaaccccaa ctttgacatc tgcaccctct tcatggaagt gatgctgccc agggaggagc 4500 4560 tetactgece ecceateace gicaaggica tegataaceg ceagtitigge egeeggeetg 4620 4680 tggtgggcca gtgtaccatc cgctccctgg agagettect gtgtgacccc tactcggcgg agagtecate cecacagggt ggeccagacg atgtgageet acteagteet ggggaagacg 4740 tgctcatcga cattgatgac aaggagcccc tcatccccat ccaggaggaa gagttcatcg 4800 attggtggag caaattcttt gcctccatag gggagaggga aaagtgcggc tcctacctgg agaaggattt tgacaccctg aaggtctatg acacacagct ggagaatgtg gaggcctttg agggcctgtc tgacttttgt aacaccttca agctgtaccg gggcaagacg caggaggaga 4860 4920 4980 5040 cagaagatcc atctgtgatt ggtgaattta agggcctctt caaaatttat cccctcccag aagacccage catecccatg cececaagae agttecacea getggeegee cagggaceee 5100 aggagtgett ggteegtate tacattgtee gageatttgg eetgeageee aaggaceeea 5160 atggaaagtg tgatccttac atcaagatct ccatagggaa gaaatcagtg agtgaccagg ataactacat ccctgcacg ctggagcccg tatttggaaa gatgttcgag ctgacctgca 5220 5280 ctctgcctct ggagaaggac ctaaagatca ctctctatga ctatgacctc ctctccaagg 5340 acgaaaagat cggtgagacg gtcgtcgacc tggagaacag gctgctgtcc aagtttgggg ctcgctgtgg actcccacag acctactgtg tctctggacc gaaccagtgg cgggaccagc tccgccctc ccagctcctc cacctcttct gccagcagca tagagtcaag gcacctgtgt 5400 5460 5520 5580 accggacaga ccgtgtaatg tttcaggata aagaatattc cattgaagag atagaggctg gcaggatece aaacecacae etgggeecag tggaggageg tetggetetg catgtgette ageageaggg cetggteceg gageaegtgg agteaeggee cetetacage eccetgeage 5640 5700 cagacatega geaggggaag etgeagatgt gggtegacet attteegaag geeetgggge ggeetggace teeetteaac ateaceceae ggagageeag aaggttttte etgegttgta 5760 5820 5880 ttatctggaa taccagagat gtgatcctgg atgacctgag cctcacgggg gagaagatga gcgacattta tgtgaaaggt tggatgattg gctttgaaga acacaagcaa aagacagacg 5940 tgcattatcg ttccctggga ggtgaaggca acttcaactg gaggttcatt ttccccttcg actacctgcc agctgagcaa gtctgtacca ttgccaagaa ggatgccttc tggaggctgg 6000 6060 acaagactga gagcaaaatc ccagcacgag tggtgttcca gatctgggac aatgacaagt 6120 teteetttga tgattttetg ggeteeetge agetegatet caacegeatg eccaagecag 6180 ccaagacage caagaagtge teettggace agetggatga tgettteeae ccagaatggt 6240 6300 ttgtgtccct ttttgagcag aaaacagtga agggctggtg gccctgtgta gcagaagagg gtgagaagaa aatactggcg ggcaagctgg aaatgacctt ggagattgta gcagagagtg 6360 agcatgagga geggeetget ggeeagggee gggatgagee caacatgaac cetaagettg 6420 aggacccaag gcgccccgac acctccttcc tgtggtttac ctccccatac aagaccatga 6480 agttcatcct gtggcggcgt ttccggtggg ccatcatcct cttcatcatc ctcttcatcc 6540 tgctgctgtt cctggccatc ttcatctacg ccttcccgaa ctatgctgcc atgaagctgg tgaagccctt cagctgagga ctctcctgcc ctgtagaagg ggccgtgggg tcccctccag 6600 6660 catgggactg gcctgcctcc tccgcccagc tcggcgagct cctccagacc tcctaggcct gattgtcctg ccagggtggg cagacagaca gatggaccgg cccacactcc cagagttgct 6720 6780 aacatggage tetgagatea ecceacttee ateattteet tetececeaa eccaacgett 6840

ttttggatca gctcagacat atttcagtat aaaacagttg gaaccacaaa aaaaaaaaa 6900 6912 aaaaaaaaa aa <210> 14 <211> 6911 <212> DNA <213> Homo sapiens <400> 14 tegacegece agecaggtge aaaatgeegt gteattggga gaeteegeag eeggageatt 60 120 agattacage tegaeggage tegggaaggg eggegggggt ggaagatgag cagaageece tgttetegga acgeeggetg acaagegggg tgagegeagg eggggegggg acceageeta 180 gcccactgga gcagccgggg gtggcccgtt cccctttaag agcaactgct ctaagccagg agccagagat tcgagccggc ctcgcccagc cagccctctc cagcgagggg acccacaagc 240 300 ggcgcetegg ceetecegae etttecgage cetettgeg ceetgggege acggggeeet 360 acacgegeca ageatgetga gggtetteat cetetatgee gagaacgtee acacaceega 420 caccgacate agegatgeet actgeteege ggtgtttgea ggggtgaaga agagaaceaa 480 agtcatcaag aacagcgtga accetgtatg gaatgaggga titgaatggg acctcaaggg catcccctg gaccaggget ctgagettca tgtggtggte aaagaccatg agacgatggg 540 600 gaggaacagg ttectggggg aagccaaggt eccacteega gaggteeteg ccaccectag 660 tetgteegee agetteaatg ecceetget ggacaccaag aageageeea caggggeete 720 780 getggteetg caggtgteet acacaceget geetggaget gtgeecetgt teeegeecee tactcetetg gagecetece egactetgee tgacetggat gtagtggeag acacaggagg 840 agaggaagac acagaggacc agggactcac tggagatgag gcggagccat tcctggatca 900 aagcggagge ccgggggete ccaccaccc aaggaaacta ccttcacgte ctecgccca 960 ctaccccggg atcaaaagaa agcgaagtgc gcctacatct agaaagctgc tgtcagacaa accgcaggat ttccagatca gggtccaggt gatcgagggg cgccagctgc cgggggtgaa 1020 1080 catcaagcet gtggtcaagg ttaccgctgc agggcagace aageggacge ggatccacaa 1140 gggaaacagc ccactettca atgagactet tttettcaac ttgtttgact eteetgggga 1200 getgtttgat gageceatet ttateaeggt ggtagaetet egttetetea ggacagatge 1260 tetecteggg gagtteegga tggaegtggg caccatttae agagageece ggeaegeeta teteaggaag tggetgetge teteagaeee tgatgaette tetgetgggg ecagaggeta 1320 1380 cctgaaaaca agcctttgtg tgctggggcc tggggacgaa gcgcctctgg agagaaaaga 1440 cccctctgaa gacaaggagg acattgaaag caacctgctc cggcccacag gcgtagccct gcgaggagcc cacttctgcc tgaaggtctt ccgggccgag gacttgccgc agatggacga 1500 1560 tgccgtgatg gacaacgtga aacagatctt tggcttcgag agtaacaaga agaacttggt ggaccccttt gtggaggtca gctttgcggg gaaaatgctg tgcagcaaga tcttggagaa gacggccaac cctcagtgga accagaacat cacactgcct gccatgtttc cctccatgtg 1620 1680 1740 cgaaaaaatg aggattegta teatagaetg ggacegeetg acteacaatg acategtgge taccacetae etgagtatgt egaaaatete tgeeeetgga ggagaaatag aagaggagee 1800 1860 tgcaggtgct gtcaagcctt cgaaagcctc agacttggat gactacctgg gcttcctccc 1920 cactitigg coctgetaca teaaceteta tggcagtece agagagtea caggettece 1980 agacccctac acagagetea acacaggeaa gggggaaggt gtggettate gtggeegget 2040 tetgetetee etggagacea agetggtgga gcacagtgaa cagaaggtgg aggacettee 2100 tgcggatgac atcctccggg tggagaagta ccttaggagg cgcaagtact ccctgtttgc ggccttctac tcagccacca tgctgtagga tgtggatgat gccatccagt ttgaggtcag 2160 2220 2280 catcgggaac tacgggaaca agttcgacat gacctgcctg ccgctggcct ccaccactca gtacagecgt geagtetttg acgggtgeca etactactae etaccetggg gtaacgtgaa 2340 acctgtggtg gtgctgtcat cctactggga ggacatcagc catagaatcg agactcagaa 2400 ccagctgctt gggattgctg accggctgga agctggcctg gagcaggtcc acctggccct 2460 gaaggegeag tgetecaegg aggaegtgga etegetggtg geteagetga eggatgaget 2520 catcgcaggc tgcagccagc ctctgggtga catccatgag acaccetctg ccacccacct 2580 ggaccagtac ctgtaccage tgegcaccca teacetgage caaateactg aggetgeeet ggecetgaag cteggccaca gtgageteee tgeagetetg gagcaggegg aggaetgget 2640 2700 cetgegtetg egtgeettgg cagaggagee ceagaacage etgeeggaca tegteatetg 2760 gatgetgeag ggagacaage gtgtggeata ceagegggtg ceegeceace aagteetett 2820 ctcccggcgg ggtgccaact actgtggcaa gaattgtggg aagctacaga caatctttct 2880 gaaatatccg atggagaagg tgcctggcgc ccggatgcca gtgcagatac gggtcaagct 2940 gtggtttggg ctctctgtgg atgagaagga gttcaaccag tttgctgagg ggaagctgtc 3000 tgtctttgct gaaacctatg agaacgagac taagttggcc cttgttggga actggggcac 3060 aacgggcctc acctaccca agttttctga cgtcacgggc aagatcaagc tacccaagga 3120 cagciticege eceteggeeg getggacetg ggetggagat tggttegtgt gteeggagaa 3180 gactetgete catgacatgg acgeeggtea cetgagette gtggaagagg tgtttgagaa 3240 ccagacccgg cttcccggag gccagtggat ctacatgagt gacaactaca ccgatgtgaa 3300 cggggagaag gtgcttccca aggatgacat tgagtgccca ctgggctgga agtgggaaga tgaggaatgg tccacagacc tcaaccgggc tgtcgatgag caaggctggg agtatagcat 3360 3420 caccatecce ceggagegga agecgaagea etgggteeet getgagaaga tgtactacae 3480

60

#### 19/68

```
acaccgacgg cggcgctggg tgcgcctgcg caggagggat ctcagccaaa tggaagcact
                                                                                                   3540
gaaaaggcac aggcaggcg aggcggaggg cgagggctgg gagtacgcct ctctttttgg ctggaagttc cacctcgagt accgcaagac agatgccttc cgccgccgcc gctggcgccg tcgcatggag ccactggaga agacgggcc tgcagctgtg tttgcccttg agggggccct
                                                                                                   3600
                                                                                                   3660
                                                                                                   3720
gggcggcgtg atggatgaca agagtgaaga ttccatgtcc gtctccacct tgagcttcgg
                                                                                                   3780
tgtgaacaga cccacgattt cctgcatatt cgactatggg aaccgctacc atctacgctg
                                                                                                    3840
                                                                                                   3900
ctacatgtac caggeceggg acctggetge gatggacaag gactetttt etgateeta
tgccatcgtc tecttectge accagageea gaagaeggtg gtggtgaaga acaceettaa
                                                                                                   3960
cccacctgg gaccagacgc tcatcttcta cgagatcgag atctttggcg agccggccac
                                                                                                   4020
                                                                                                   4080
agttgctgag caaccgccca gcattgtggt ggagctgtac gaccatgaca cttatggtgc
agacgagttt atgggteget geatetgtea accgagtetg gaacggatge caeggetgge
                                                                                                   4140
ctggttccca ctgacgaggg gcagccagcc gtcgggggag ctgctggcct cttttgagct catccagaga gagaagccgg ccatccacca tattcctggt tttgaggtgc aggagacatc
                                                                                                   4200
                                                                                                    4260
aaggateetg gatgagtetg aggacacaga cetgeeetae ceaceacee agagggagge caacatetae atggtteete agaacateaa gecagegete cagegtaceg ceategagat
                                                                                                    4320
                                                                                                   4380
cctggcatgg ggcctgcgga acatgaagag ttaccagctg gccaacatct cctccccag cctcgtggta gagtgtgggg gccagacggt gcagtcctgt gtcatcagga acctccggaa
                                                                                                    4440
                                                                                                    4500
gaaccccaac tttgacatct gcaccctctt catggaagtg atgctgccca gggaggagct ctactgcccc cccatcaccg tcaaggtcat cgataaccgc cagtttggcc gccggcctgt
                                                                                                    4560
                                                                                                    4620
ggtgggccag tgtaccatcc gctccctgga gagcttcctg tgtgacccct actcggcgga
                                                                                                    4680
                                                                                                    4740
gagtecatec ccacagggtg geccagaega tgtgageeta etcagteetg gggaagaegt
                                                                                                    4800
gctcatcgac attgatgaca aggagcccct catccccatc caggaggaag agttcatcga
                                                                                                    4860
ttggtggagc aaattetttg cetecatagg ggagagggaa aagtgegget eetacetgga
gaaggatttt gacaccctga aggtctatga cacacagctg gagaatgtgg aggcctttga
                                                                                                    4920
gggcctgtct gacttttgta acaccttcaa gctgtaccgg ggcaagacgc aggaggagac agaagatcca tctgtgattg gtgaatttaa gggcctcttc aaaatttatc ccctcccaga
                                                                                                    4980
                                                                                                    5040
agacccagec atececatge ecccaagaca gttecaccag etggeegeec agggaececa
                                                                                                    5100
ggagtgcttg gtccgtatct acattgtccg agcatttggc ctgcagccca aggaccccaa tggaaagtgt gatccttaca tcaagatctc catagggaag aaatcagtga gtgaccagga taactacatc ccctgcacgc tggagcccgt atttggaaag atgttcgagc tgacctgcac
                                                                                                    5160
                                                                                                    5220
                                                                                                    5280
tetgeetetg gagaaggaee taaagateae tetetatgae tatgaeetee tetecaagga
                                                                                                    5340
cgaaaagatc ggtgagacgg tcgtcgacct ggagaacagg ctgctgtcca agtttggggc
                                                                                                    5400
tcgctgtgga ctcccacaga cctactgtgt ctctggaccg aaccagtggc gggaccagct ccgccctcc cagctcctcc acctcttctg ccagcagcat agagtcaagg cacctgtgta ccggacagac cgtgtaatgt ttcaggataa agaatattcc attgaagaga tagaggctgg
                                                                                                    5460
                                                                                                    5520
                                                                                                    5580
 caggatecca aacccacace tgggcccagt ggaggagegt etggetetge atgtgettea
                                                                                                    5640
 gcagcagggc ctggtcccgg agcacgtgga gtcacggccc ctctacagcc ccctgcagcc
                                                                                                    5700
                                                                                                    5760
 agacatcgag caggggaagc tgcagatgtg ggtcgaccta tttccgaagg ccctggggcg
gcctggacct cccttcaaca tcaccccacg gagagccaga aggtttttcc tgcgttgtat tatctggaat accagagatg tgatcctgga tgacctgagc ctcacggggg agaagatgag cgacatttat gtgaaaggtt ggatgattgg ctttgaagaa cacaagcaaa agacagacgt
                                                                                                    5820
                                                                                                    5880
                                                                                                    5940
 gcattategt tecetgggag gtgaaggeaa etteaactgg aggtteattt teecettega
                                                                                                    6000
                                                                                                    6060
 ctacctgcca gctgagcaag tctgtaccat tgccaagaag gatgccttct ggaggctgga
 caagactgag agcaaaatcc cagcacgagt ggtgttccag atctgggaca atgacaagtt
                                                                                                    6120
 ctcctttgat gattttctgg gctccctgca gctcgatctc aaccgcatgc ccaagccagc caagacagcc aagaagtgct ccttggacca gctggatgat gctttccacc cagaatggtt
                                                                                                    6180
                                                                                                    6240
 tgtgtccctt tttgagcaga aaacagtgaa gggctggtgg ccctgtgtag cagaagaggg
                                                                                                    6300
 tgagaagaaa atactggcgg gcaagctgga aatgaccttg gagattgtag cagagagtga gcatgagga cggcctgctg gccagggccg ggatgagccc aacatgaacc ctaagcttga ggacccaagg cgccccgaca cctccttcct gtggtttacc tccccataca agaccatgaa gttcatcctt tggcggcgtt tccggtgggc catcatcctc ttcatcatcct tcttcatcct
                                                                                                    6360
                                                                                                    6420
                                                                                                    6480
                                                                                                    6540
                                                                                                    6600
 getgetgtte etggecatet teatetacge ettecegaae tatgetgeca tgaagetggt
 gaagecette agetgaggae teteetgeee tgtagaaggg geegtggggt cecetecage atgggaetgg cetgeeteet cegeceaget eggegagete etceagacet ectaggeetg
                                                                                                    6660
                                                                                                    6720
                                                                                                    6780
 attgtcctgc cagggtgggc agacagacag atggaccggc ccacactece agagttgcta
 acatggaget etgagateac eccaetteca teattteett eteccecaac ecaacgettt
                                                                                                    6840
 tttggatcag ctcagacata tttcagtata aaacagttgg aaccacaaaa aaaaaaaaa
                                                                                                    6900
                                                                                                    6911
 aaaaaaaaa a
```

<210> 15 <211> 6910

<212> DNA <213> Homo sapiens

<400> 15

tegacegece agecaggtge aaaatgeegt gteattggga gaeteegeag eeggageatt agattacage tegacggage tegggaaggg eggegggggt ggaagatgag cagaageeee 120

					_	
tattctcaga	acgccggctg	acaagcgggg	tgagcgcagg	cggggcgggg	acccagccta	180
accasataas	gcagccgggg	atageceatt	cccctttaag	aggaagtggt	ctaagccagg	240
geceaetgga	gcagccgggg	gradeceare	ccccccaag	ageadorgee	ccaagccagg	
agccagagat	tcgagccggc	ctcgcccagc	cagccctctc	cagcgagggg	acccacaagc	300
agagast can	ccctcccgac	ctttccgage	cctctttaca	ccctagacac	acqqqqcct	360
ggegeeeegg	ccccccgac	ccccogage				
acacgcgcca	agcatgctga	gggtcttcat	cctctatgcc	gagaacgtcc	acacacccga	420
caccoacate	agcgatgcct	actoctccoc	ggtgtttgca	ggggtgaaga	agagaaccaa	480
cacegacace	agogaogeo		55-55	5555-55-		
agtcatcaag	aacagcgtga	accctgtatg	gaatgaggga	rrrgaarggg	acctcaaggg	540
ant accept a	gaccagggct	ctgaggttca	tataataata	aaagaccatg	agacgatggg	600
Cattlettettg	gaccagggcc	ccgageeeea	cgcggcggcc	adagaeeaeg	agacgacggg	
gaggaacagg	ttcctggggg	aagccaaggt	cccactccga	gaggtcctcg	ccacccctag	660
***	agcttcaatg	acceptact	adacaccaaa	aaggagggga	caggggggtt	720
tetgteegee	agetteaatg	ccccccgcc	ggacaccaag	augeugeeeu	caggggcccc	
actaatccta	caggtgtcct	acacaccgct	gcctggagct	gtgcccctgt	tcccgcccc	780
++	gagecetece	agastataca	taacctaat	atentageen	acacadada	840
tactectetg	gageeeeee	cgactetgee	cgacccggac	gragrageag	acacaggagg	
agaggaagac	acagaggacc	agggactcac	tqqaqatqaq	gcggagccat	tcctggatca	900
-9-999	accesses to	222	22772227	cetterecate	ctcccccca	960
aageggagge	ccgggggctc	CCaccacccc	aayyaaacta	ccccacqcc	cccgcccca	
ctaccccaaa	atcaaaagaa	agcgaagtgc	gcctacatct	agaaagctgc	tgtcagacaa	1020
	ttccagatca	agatagaagt	astensagaa	caccaactac	caaaaataaa	1080
accycaygat	LLCCagacca	gggtccaggt	gaccgagggg	cgccagccgc	cgggggcgaa	_
catcaagect	gtggtcaagg	ttaccqctqc	agggcagacc	aagcggacgc	ggatccacaa	1140
		ataaaaatat	tttattanna	ttatttaact	at cot again	1200
gggaaacagc	ccactcttca	atgagactet	CCCCCCCaac	Ligitigati	ccccgggga	
actatttaat	gagcccatct	ttatcacggt	ggtagactct	cattctctca	ggacagatgc	1260
googoogae	949000000				~~~~~	1320
tctcctcggg	gagttccgga	tggacgtggg	Caccatttac	agagageeee	ggcacgccca	
teteaggaag	tggctgctgc	tctcagaccc	tgatgacttc	tctactaaaa	ccagaggcta	1380
cocoaggaag	-9999-					1440
cctgaaaaca	agcctttgtg	tgctggggcc	tggggacgaa	gegeetetgg	agagaaaaga	1440
ccctctcaa	gacaaggagg	acattgaaag	caacctgctc	cggcccacag	gcgtagccct	1500
ccccccgaa	gacaaggagg	dodoogaaag	on a constant	-99	909229000	
gcgaggagcc	cacttctgcc	tgaaggtett	ccgggccgag	gacttgccgc	agatggacga	1560
+	gacaacgtga	aacacatctt	taacttcaaa	antaacaaca	agaacttagt	1620
egeegegatg	gacaacgiga	aacagacccc	cggccccgag	agcaacaaga	agaacccggc	
ggaccccttt	gtggaggtca	gctttgcggg	gaaaatgctg	tgcagcaaga	tcttggagaa	1680
~~~~~	antanatana	200202202	cacactacct	accetattta	cctccatata	1740
gaeggeeaac	cctcagtgga	accagaacac	Cacaccgccc	gccacgcccc	ccccacgcg	
cgaaaaaatg	aggattcgta	tcatagactg	ggaccgcctg	actcacaatq	acatcqtqqc	1800
+	ctgagtatgt	agaaatete	tacccctaca	gracestag	22020200	1860
taccacctac	Cigagiaigi	Cyaaaacccc	egeceeegga	ggagaaacag	aagaggagcc	
tacagatact	gtcaagcctt	cgaaagcctc	agacttggat	gactacctgg	gcttcctccc	1920
	ccctgctaca	+02200+0+2	taggaatagg	agagagttca	caggetteec	1980
agacccctac	acagagctca	acacaggcaa	gggggaaggt	gtggcttatc	gtggccggct	2040
+ =+ ==+ =+ ==	ctggagacca	agetagtage	acacaataaa	cadaadataa	aggacettee	2100
Leigerere	Ciggagacca	ageegge	gcacagcgaa	cagaaggegg	aggaceeeee	
tacagataac	atcctccggg	tggagaagta	ccttaggagg	cgcaagtact	ccctgtttgc	2160
accettat ac	tcagccacca	+actacaaaa	tatagataat	accatecant	ttgaggtgag	2220
ggeettetae	ccagccacca	Lyceycayya	cgcggacgac	gecaectage	cegaggeeag	
catcgggaac	tacgggaaca	agttcgacat	gacctgcctg	ccqctqqcct	ccaccactca	2280
	gcagtctttg	agagt acca	ctactactac	ctaccetaca	ataacataaa	2340
gracageegr	geageeeeg	acgggcgcca	CCaccaccac	ccaccccggg	gedaegegaa	
acctataata	gtgctgtcat	cctactqqqa	ggacatcagc	catagaatcg	agactcagaa	2400
	~~~~++~~+~	acconstant	agetggggtg	gaggagates	acctagacact	2460
ecagetgett	gggattgctg	accygccyga	ageeggeeeg	gageaggeee	acceggeeee	
gaaggcgcag	tgctccacgg	aggacgtgga	ctcqctqqtq	gctcagctga	cggatgagct	2520
9		atataaataa	ant agat and	agaggt at a	aaaaaaaat	2580
categeagge	tgcagccagc	ccccgggcga	catcuatgag	acaccccccg	CCacccaccc	
ggaccagtac	ctgtaccagc	tacacaccca	tcacctgage	caaatcactq	aggetgeect	2640
		at as act as a	taasaatata	a2aa2aaaa	aga at agat	2700
ggccctgaag	ctcggccaca	grgagereee	tgcagctctg	gagcaggcgg	aggactggct	
cctacatcta	cgtgccctgg	cagaggagcc	ccaqaacaqc	ctgccggaca	tcgtcatctg	2760
	-5-555	-+-+	aaaaaaata	accacaca	220+00+0++	2820
gatgetgeag	ggagacaagc	gigiggiala	ccagegggtg	deegeccade	aagttttt	
ctcccaacaa	ggtgccaact	actotoocaa	gaattgtggg	aagctacaga	caatctttct	2880
		taaataaaa	accastaccs	atacagatac	aaataaaaat	2940
gaaatateeg	atggagaagg	egeeeggege	ccggatgcca	gracagarac	gggccaagec	
ataatttaaa	ctctctgtgg	atgagaagga	gttcaaccag	tttactaaaa	ggaagctgtc	3000
3-35-C5339			taaattaa==	a++a++~~~	201000000	3060
tgtctttgct	gaaacctatg	agaacgagac	LaagLLggcc	cttgttggga	actggggeac	
aacqqqcctc	acctacccca	agttttctga	catcacaaac	aagatcaagc	tacccaagga	3120
			~~~	++	-+	3180
cagetteege	ccctcggccg	getggaeetg	ggctggagat	rggrregrgr	greeggagaa	
gactetgete	catgacatgg	acaccaatca	cctgagcttc	gtggaagagg	tatttaagaa	3240
gastatgas				3-33-3-33		3300
ccagacccgg	cttcccggag	gecageggae	CLacatgagt	gacaactaca	cegatgtgaa	
cggggagaag	gtgcttccca	aggatgacat	tgagtgccca	ctqqqctqqa	agtgggaaga	3360
	± c c c c c c c c c c c c c c c c c c c	+ 02 2000000	tataastasa	~==~~~	agtataggat	3420
tgaggaatgg	tccacagacc	ccaaccygyc	cyccyacyag	caayyorgyg	aycacaycat	
caccatecee	ccggagcgga	agccgaagca	ctgggtccct	gctgagaaga	tgtactacac	3480
		+~~~~	0200200	otosoccc-	+00000000±	3540
	cggcgctggg					
gaaaagggag	aggcaggcgg	aggcagagag	cgagggctgg	gagtaccct	ctctttttaa	3600
		22-22-22	202535	00000000	antacasas	3660
ctggaagttc	cacctcgagt	accycaagac	agargeette	uguagaagaa	gerggegeeg	
tegeategag	ccactggaga	agacggggcc	tgcagctgtg	tttqccctta	agggggccct	3720
gggcggcgtg	atggatgaca	agagtgaaga	LECCATGECC	grorecacet	rgagettegg	3780
tataaacaaa	cccacgattt	cctccatatt	cgactatggg	aaccoctacc	atctacocto	3840
-g-g-c-ga			5			
ctacatgtac	caggcccggg	acctggctgc	gatggacaag	gactctttt	ctgatcccta	3900
taccateata	tccttcctgc	accadadeca	gaagacggtg	ataataaaaa	acacccttaa	3960
Lycoallyco		trade to the		2-22-24uga		
ccccacctgq	gaccagacgc	tcatcttcta	cgagatcgag	atctttggcg	agccggccac	4020
agttggtgag	caaccgccca	acattataat	ggagctgtac	gaccatgaca	cttatootoc	4080
-googe coug			2000000000	2222222		
agacgagttt	atgggtcgct	gcatctgtca	accgagtetg	gaacggatgc	cacggctggc	4140

```
ctggttccca ctgacgaggg gcagccagcc gtcgggggag ctgctggcct cttttgagct
                                                                                       4200
catccagaga gagaagccgg ccatccacca tattcctggt tttgaggtgc aggagacatc
                                                                                       4260
aaggateetg gatgagtetg aggacacaga cetgecetae ecaceacece agagggagge
                                                                                       4320
caacatctac atggttcctc agaacatcaa gccagcgctc cagcgtaccg ccatcgagat
                                                                                       4380
cctggcatgg ggcctgcgga acatgaagag ttaccagctg gccaacatct cctccccag cctcgtggta gagtgtgggg gccagacggt gcagtcctgt gtcatcagga acctccggaa gaaccccaac tttgacatct gcaccctctt catggaagtg atgctgccca gggaggagct
                                                                                       4440
                                                                                       4500
                                                                                       4560
                                                                                       4620
ctactgccc cccatcaccg tcaaggtcat cgataaccgc cagtttggcc gccggcctgt
ggtgggccag tgtaccatcc gctccctgga gagcttcctg tgtgacccct actcggcgga gagtccatcc ccacagggtg gcccagacga tgtgagccta ctcagtcctg gggaagacgt gctcatcgac attgatgaca aggagcccct catccccatc caggaggaag agttcatcga
                                                                                       4680
                                                                                       4740
                                                                                       4800
ttggtggagc aaattotttg cotocatagg ggagagggaa aagtgcggot cotacotgga
                                                                                       4860
                                                                                       4920
quaggattit gacaccctga aggtetatga cacacagetg gagaatgtgg aggeetttga
gggcctgtct gacttttgta acaccttcaa gctgtaccgg ggcaagacgc aggaggagac agaagatcca tctgtgattg gtgaatttaa gggcctcttc aaaatttatc ccctcccaga
                                                                                       4980
                                                                                       5040
agacccagcc atccccatgc ccccaagaca gttccaccag ctggccgccc agggacccca
                                                                                       5100
                                                                                       5160
ggagtgcttg gtccgtatct acattgtccg agcatttggc ctgcagccca aggaccccaa
                                                                                       5220
tggaaagtgt gatccttaca tcaagatctc catagggaag aaatcagtga gtgaccagga
taactacate ceetgeacge tggageeegt atttggaaag atgttegage tgacetgeac
                                                                                       5280
tetgeetetg gagaaggace taaagateae tetetatgae tatgaeetee tetecaagga egaaaagate ggtgagaegg tegtegaeet ggagaacagg etgetgteea agtttgggge
                                                                                       5340
                                                                                       5400
                                                                                       5460
tegetgtgga eteceacaga cetaetgtgt etetggaceg aaccagtgge gggaccaget
cogcocctcc cagetectcc acetettetg ccageageat agagteaagg cacetgtgta
                                                                                       5520
ccggacagac cgtgtaatgt ttcaggataa agaatattcc attgaagaga tagaggctgg
                                                                                       5580
caggatecca aacceacace tgggeccagt ggaggagegt etggetetge atgtgettea geageaggge etggteegg ageaegtgga gteaeggeee etetaeagee eeetgeagee
                                                                                       5640
                                                                                       5700
agacategag caggggaage tgcagatgtg ggtcgaceta tttccgaagg ccctggggcg
                                                                                       5760
gcctggacct cccttcaaca tcaccccacg gagagccaga aggtttttcc tgcgttgtat
                                                                                       5820
tatctggaat accagagatg tgatcctgga tgacctgagc ctcacggggg agaagatgag
                                                                                       5880
                                                                                       5940
cgacatttat gtgaaaggtt ggatgattgg ctttgaagaa cacaagcaaa agacagacgt gcattatcgt tccctgggag gtgaagcaac ttcaactgga ggttcatttt ccccttcgac
                                                                                       6000
tacctgccag ctgagcaagt ctgtaccatt gccaagaagg atgccttctg gaggctggac
                                                                                       6060
aagactgaga gcaaaatccc agcacgagtg gtgttccaga tctgggacaa tgacaagttc
                                                                                       6120
tcctttgatg attttctggg ctccctgcag ctcgatctca accgcatgcc caagccagcc aagacagcca agaagtgctc cttggaccag ctggatgatg ctttccaccc agaatggttt
                                                                                       6180
                                                                                       6240
gtgtcccttt ttgagcagaa aacagtgaag ggctggtggc cctgtgtagc agaagagggt
                                                                                       6300
                                                                                       6360
gagaagaaaa tactggcggg caagctggaa atgaccttgg agattgtagc agagagtgag
catgaggage ggeetgetgg ecagggeegg gatgageeca acatgaacce taagettgag
                                                                                       6420
gacccaagge geecegacae etectteetg tggtttacet ecccatacaa gaccatgaag
                                                                                       6480
ttcatcctgt ggcggcgttt ccggtgggcc atcatcctct tcatcatcct cttcatcctg
                                                                                       6540
ctgctgttcc tggccatctt catctacgcc ttcccgaact atgctgccat gaagctggtg
                                                                                       6600
aagcccttca gctgaggact ctcctgccct gtagaagggg ccgtggggtc ccctccagca
                                                                                       6660
tgggactggc ctgcctcctc cgcccagctc ggcgagctcc tccagacctc ctaggcctga
                                                                                       6720
ttgtcctgcc agggtgggca gacagacaga tggaccggcc cacactccca gagttgctaa
                                                                                       6780
catggagete tgagateace ceaetteeat cattteette tececeaace caaegetttt ttggateage teagacatat tteagtataa aaeagttgga accacaaaaa aaaaaaaaa
                                                                                       6840
                                                                                       6900
                                                                                       6910
aaaaaaaaa
         <210> 16
         <211> 6911
         <212> DNA
         <213> Homo sapiens
         <400> 16
                                                                                          60
togacogoco agocaggtgo aaaatgoogt gtoattggga gactoogoag coggagoatt
                                                                                         120
```

agattacage tegacggage tegggaaggg eggegggggt ggaagatgag cagaageeee 180 gcccactgga gcagccgggg gtggcccgtt cccctttaag agcaactgct ctaagccagg agccagagat tcgagccggc ctcgcccagc cagccctctc cagcgagggg acccacaagc ggcgcctcgg ccctcccgac ctttccgagc cctctttgcg ccctgggcgc acggggccct 240 300 360 acacgogoca agcatgotga gggtottoat cototatgoo gagaacgtoo acacaccoga 420 480 caccgacate agegatgeet actgeteege ggtgtttgca ggggtgaaga agagaaccaa agtcatcaag aacagcgtga accetgtatg gaatgaggga tttgaatggg accetaaggg catcccctg gaccaggget ctgagettea tgtggtggte aaagaccatg agacgatggg 540 600 gaggaacagg ttoctggggg aagccaaggt cccactccga gaggtcctcg ccaccctag tctgtccgcc agcttcaatg ccccctgct ggacaccaag aagcagccca caggggcctc 660 720 780 getggteetg caggtgteet acacaceget geetggaget gtgeecetgt teeegeeeee

tactcctctg gagccctccc cgactctgcc tgacctggat gtagtggcag acacaggagg 840 agaggaagac acagaggacc agggactcac tggagatgag gcggagccat tcctggatca 900 aagcggagge ccgggggete ccaccaccc aaggaaacta cettcacgte ctccgccca 960 ctacccggg atcaaaagaa agcgaagtgc gcctacatct agaaagctgc tgtcagacaa 1020 accgcaggat ttccagatca gggtccaggt gatcgagggg cgccagctgc cgggggtgaa 1080 catcaagcct gtggtcaagg ttaccgctgc agggcagacc aagcggacgc ggatccacaa 1140 gggaaacage ceactettea atgagactet titetteaac ttgtitgaet etectgggga 1200 gctgtttgat gagcccatct ttatcacggt ggtagactct cgttctctca ggacagatgc 1260 tetecteggg gagtteegga tggacgtggg caccatttae agagageece ggeacgeeta teteaggaag tggetgetge teteagacee tgatgaette tetgetgggg ceagaggeta 1320 1380 cctgaaaaca agcctttgtg tgctggggcc tggggacgaa gcgcctctgg agagaaaaga cccctctgaa gacaaggagg acattgaaag caacctgctc cggcccacag gcgtagccct 1440 1500 gcgaggagcc cacttetgee tgaaggtett ccgggeegag gaettgeege agatggaega tgeegtgatg gaeaacgtga aacagatett tggettegag agtaacaaga agaacttggt 1560 1620 ggaccccttt gtggaggtca gctttgcggg gaaaatgctg tgcagcaaga tcttggagaa gacggccaac cctcagtgga accagaacat cacactgcct gccatgtttc cctccatgtg 1680 1740 cgaaaaaatg aggattegta teatagaetg ggaeegeetg acteacaatg acategtgge taccaectae etgagtatgt egaaaatete tgeeeetgga ggagaaatag aagaggagee 1800 1860 tgcaggtgct gtcaagcctt cgaaagcctc agacttggat gactacctgg gcttcctccc 1920 cacttitggg ccctgctaca tcaacctcta tggcagtccc agagagttca caggettccc 1980 agacccctac acagagetea acacaggeaa gggggaaggt gtggettate gtggeegget 2040 2100 tetgetetee etggagacca agetggtgga geacagtgaa cagaaggtgg aggacettee tgcggatgac atcctccggg tggagaagta ccttaggagg cgcaagtact ccctgtttgc 2160 ggccttctac tcagccacca tgctgcagga tgtggatgat gccatccagt ttgaggtcag 2220 categggaac tacgggaaca agttegacat gacetgeetg cegetggeet ceaceactea gtacageegt geagtetttg aegggtgeea etactactae etaceetggg gtaaegtgaa 2280 2340 acctgtggtg gtgctgtcat cctactggga ggacatcagc catagaatcg agactcagaa ccagctgctt gggattgctg accggctgga agctggcctg gagcaggtcc acctggccct 2400 2460 gaaggegeag tgeteeaegg aggaegtgga etegetggtg geteagetga eggatgaget categeagge tgeageeage etetgggtga categatgag acaccetetg ceacceaect 2520 2580 ggaccagtac ctgtaccage tgegcaccca teacetgage caaateactg aggetgeeet 2640 ggccctgaag ctcggccaca gtgagctccc tgcagctctg gagcaggcgg aggactggct 2700 cetgegtetg cgtgecetgg cagaggagec ccagaacage etgeeggaca tegteatetg 2760 gatgetgeag ggagacaage gtgtggeata ceagegggtg ceegeceace aagteetett eteeggegg ggtgeeaact actgtggeaa gaattgtggg aagetacaga caatettet 2820 2880 gaaatatccg atggagaagg tgcctggcgc ccggatgcca gtgcagatac gggtcaagct gtggtttggg ctctctgtgg atgagaagga gttcaaccag tttgctgagg ggaagctgtc tgtctttgct gaaacctatg agaacgagac taagttggcc cttgttggga actggggcac 2940 3000 3060 aacgggcete acctacecca agttttetga egteaeggge aagateaage tacccaagga 3120 cagettecge eceteggeeg getggacetg ggetggagat tggttegtgt gteeggagaa gaetetgete catgacatgg acgeeggtea ectgagette gtggaagagg tgtttgagaa 3180 3240 ccagaccegg cttcccggag gccagtggat ctacatgagt gacaactaca ccgatgtgaa 3300 cggggagaag gtgcttcca aggatgacat tgagtgcca ctgggctgga agtgggaaga 3360 tgaggaatgg tccacagacc tcaaccgggc tgtcgatgag caaggctggg agtatagcat 3420 caccatccc ccggagcgga agccgaagca ctgggtccct gctgagaaga tgtactacac 3480 acaccgacgg cggcgctggg tgcgcctgcg caggagggat ctcagccaaa tggaagcact gaaaaggcac aggcaggcgg aggcggaggg cgagggctgg gagtacgcct ctctttttgg ctggaagttc cacctcgagt accgcaagac agatgccttc cgccgccgcc gctggcgccg 3540 3600 3660 tegeatggag ceaetggaga agaeggggee tgeagetgtg titgeeettg agggggeeet gggeggegtg atggatgaca agagtgaaga tteeatgtee gtetecaeet tgagettegg 3720 3780 tgtgaacaga cccacgattt cctgcatatt cgactatggg aaccgctacc atctacgctg 3840 ctacatgtac caggeceggg acctggetge gatggacaag gactetttt etgateceta 3900 tgccatcgte tecttectge accagageea gaagaeggtg gtggtgaaga acaceettaa ecceacetgg gaccagaege teatetteta egagategag atetttggeg ageeggeeac 3960 4020 agttgctgag caaccgccca gcattgtggt ggagctgtac gaccatgaca cttatggtgc agacgagttt atgggtcgct gcatctgtca accgagtctg gaacggatgc cacggctggc 4080 4140 ctggttccca ctgacgaggg gcagccagec gtcgggggag ctgctggcct cttttgagct 4200 catccagaga gagaagccgg ccatccacca tattcctggt tttgaggtgc aggagacatc 4260 aaggateetg gatgagtetg aggacacaga cetgecetae ceaceaeece agagggagge caacatetae atggtteete agaacateaa gecagegete cagegtaeeg ceategagat 4320 4380 cctggcatgg ggcctgcgga acatgaagag ttaccagctg gccaacatct cctccccag 4440 cctcgtggta gagtgtgggg gccagacggt gcagtcctgt gtcatcagga acctccggaa 4500 4560 gaaccccaac tttgacatct gcaccctctt catggaagtg atgctgccca gggaggagct ctactgcccc cccatcaccg tcaaggtcat cgataaccgc cagtttggcc gccggcctgt 4620 ggtgggccag tgtaccatcc gctccctgga gagcttcctg tgtgacccct actcggcgga 4680 4740 gagtecatec ccacagggtg geccagaega tgtgageeta etcagteetg gggaagaegt getcategae attgatgaca aggageceet catececate caggaggaag agttcatega 4800

ttaataaaa	aaattctttg	cctccatagg	ggagaggaa	aagtgcggct	cctacctgga	4860
gaaggatttt	gacaccctga	aggtctatga	cacacagetg		aggcctttga	4920
gaaggactatet	gacttttgta	acaccttcaa	gctgtaccgg	ggcaagacgc		4980
agaagatcca	tctgtgattg	gtgaatttaa	aggeetette	• • •	ccctcccaga	5040
agaagacca	atccccatgc	ccccaagaca	attccaccag	ctggccgccc	_	5100
agactcage	gtccgtatct	acattotcco	agcatttggc	ctgcagccca		5160
tagaaagtat	gatccttaca	tcaagatctc	catagggaag	aaatcagtga	gtgaccagga	5220
taactacatc	ccctgcacgc	tagaacccat	atttggaaag		tgacctgcac	5280
tetaceteta	gagaaggacc	taaagatcac	tctctatqac	tatgacctcc	tctccaagga	5340
cgaaaagatc	ggtgagacgg	tcqtcqacct	ggagaacagg	ctgctgtcca	agtttggggc	5400
tcactataga	ctcccacaga	cctactatat	ctctggaccg	aaccagtggc	gggaccagct	5460
ccaccactca	cagctcctcc	acctcttctg	ccaqcaqcat	agagtcaagg	cacctgtgta	5520
ccgacagac	cgtgtaatgt	ttcaggataa	agaatattcc		tagaggctgg	5580
caggatecca	aacccacacc	tagacccagt	ggaggagcgt			5640
acaacaaaac	ctggtcccgg	agcacqtqqa	gtcacggccc	ctctacagcc	ccctgcagcc	5700
agacatcgag	caggggaagc	tgcagatgtg	ggtcgaccta	tttccgaagg	ccctggggcg	5760
acctagacct	cccttcaaca	tcaccccacg	gagagccaga	aggtttttcc	tgcgttgtat	5820
tatctggaat	accagagatg	tgatcctgga	tgacctgagc	ctcacggggt	agaagatgag	5880
	gtgaaaggtt	ggatgattgg	ctttgaagaa	cacaagcaaa	agacagacgt	5940
	tccctgggag	gtgaaggcaa	cttcaactgg	aggttcattt	tccccttcga	6000
	gctgagcaag	tctgtaccat	tgccaagaag	gatgccttct	ggaggctgga	6060
	agcaaaatcc	cagcacgagt	ggtgttccag	atctgggaca	atgacaagtt	6120
	gattttctgg		gctcgatctc	aaccgcatgc	ccaagccagc	6180
	aagaagtgct	T	gctggatgat	gctttccacc	cagaatggtt	6240
	tttgagcaga		gggctggtgg	ccctgtgtag	cagaagaggg	6300
tgagaagaaa		gcaagctgga	aatgaccttg	gagattgtag	cagagagtga	6360
qcatqaggag	cagcetgetg	gccagggccg	ggatgagccc	aacatgaacc	ctaagcttga	6420
ggacccaagg	cgccccgaca	cctccttcct	gtggtttacc	tccccataca	agaccatgaa	6480
gttcatcctg	tggcggcgtt	tccggtgggc	catcatcctc	ttcatcatcc	tcttcatcct	6540
actactatte	ctggccatct	tcatctacgc	cttcccgaac	tatgctgcca	tgaagctggt	6600
gaagcccttc	agctgaggac	tetectgece	tgtagaaggg	gccgtggggt	cccctccagc	6660
atgggactgg	cctqcctcct	ccgcccagct	cggcgagctc	ctccagacct	cctaggcctg	6720
attotectoe	cagggtgggc	agacagacag	atggaccggc	ccacactccc	agagttgcta	6780
acatggagct	ctgagatcac	cccacttcca	tcatttcctt	ctccccaac	ccaacgcttt	6840
tttggatcag	ctcagacata	tttcagtata	aaacagttgg	aaccacaaaa	aaaaaaaaa	6900
aaaaaaaaa		_				6911

<210> 17 <211> 6911 <212> DNA <213> Homo sapiens

#### <400> 17

60 tcgaccgccc agccaggtgc aaaatgccgt gtcattggga gactccgcag ccggagcatt agattacage tegacggage tegggaaggg eggegggggt ggaagatgag eagaageeee 120 180 gcccactgga gcagccgggg gtggcccgtt cccctttaag agcaactgct ctaagccagg 240 agccagagat togagooggo otogoccago cagocototo cagogagggg acccacaago 300 ggcgcctcgg ccctcccgac ctttccgagc cctctttgcg ccctgggcgc acggggccct 360 acacgogoca agoatgotga gggtottoat cototatgoo gagaacgtoo acacaccoga 420 caccgacate agegatgeet actgeteege ggtgtttgea ggggtgaaga agagaaccaa 480 540 agtcatcaag aacagcgtga accetgtatg gaatgaggga tttgaatggg acctcaaggg catcccctg gaccagggct ctgagcttca tgtggtggtc aaagaccatg agacgatggg 600 gaggaacagg ttcctggggg aagccaaggt cccactccga gaggtcctcg ccacccctag tctgtccgc agcttcaatg ccccctgct ggacaccaag aagcagccca caggggcctc gctggtcctg caggtgtcct acacaccgct gcctggagct gtgcccctgt tcccgccccc 660 720 780 tactcctctg gagccctccc cgactctgcc tgacctggat gtagtggcag acacaggagg 840 900 agaggaagac acagaggacc agggactcac tggagatgag gcggagccat tcctggatca 960 aagcggagge cegggggete ceaceacece aaggaaacta cetteacgte etecgececa ctaccccggg atcaaaagaa agcgaagtgc gcctacatct agaaagctgc tgtcagacaa 1020 accgcaggat ttccagatca gggtccaggt gatcgagggg cgccagctgc cgggggtgaa 1080 1140 catcaageet gtggtcaagg ttaccgetge agggcagace aageggacge ggatecacaa gggaaacage ccactettea atgagaetet titetteaae ttgtitgaet eteetgggga 1200 1260 gctgtttgat gagcccatct ttatcacggt ggtagactct cgttctctca ggacagatgc tctcctcggg gagttccgga tggacgtggg caccatttac agagagcccc ggcacgccta tctcaggaag tggctgctgc tctcagaccc tgatgacttc tctgctgggg ccagaggcta 1320 1380 cctgaaaaca agcctttgtg tgctggggcc tggggacgaa gcgcctctgg agagaaaaga 1440 ccctctgaa gacaaggagg acattgaaag caacctgctc cggcccacag gcgtagccct 1500 gcgaggagcc cacttetgcc tgaaggtett ccgggccgag gacttgccgc agatggacga 1560 tgccgtgatg gacaacgtga aacagatctt tggcttcgag agtaacaaga agaacttggt 1620 ggaccecttt gtggaggtca getttgeggg gaaaatgetg tgeageaaga tettggagaa gaeggeeaac ceteagtgga accagaacat cacactgeet geeatgttte cetecatgtg 1680 1740 cgaaaaaatg aggattegta teatagactg ggacegeetg acteacaatg acategtgge 1800 taccacctac ctgagtatgt cgaaaatctc tgcccctgga ggagaaatag aagaggagcc 1860 tgcaggtgct gtcaagcctt cgaaagcctc agacttggat gactacctgg gcttcctccc 1920 cacttttggg ccctgctaca tcaacctcta tggcagtccc agagagttca caggcttccc 1980 agacccctac acagagetea acacaggeaa gggggaaggt gtggettate gtggeegget 2040 tetgetetec etggagacca agetggtgga gcacagtgaa cagaaggtgg aggacettee 2100 tgcggatgac atcctccggg tggagaagta ccttaggagg cgcaagtact ccctgtttgc 2160 ggccttctac tcagccacca tgctgcagga tgtggatgat gccatccagt ttgaggtcag 2220 catcgggaac tacgggaaca agttcgacat gacctgcctg ccgctggcct ccaccactca 2280 gtacagcegt gcagtetttg acgggtgcca ctactactac ctaccetggg gtaacgtgaa acctgtggtg gtgetgtcat cctactgga ggacatcagc catagaatcg ggattgctg accggctgga agctggcctg gagcaggtcc acctggccct gaaggcgcag tgctccacgg aggacgtgga catcgtggg gctcagctga cggatgagct tgcagcagc ctctgggtga catcatgag acaccetctg caaccacct ggagcagtag ctctgggtga catcatgag acaccetctg ccaccacct ggagcagtag ctctgggtga catcatgag acaccetctg ccaccacct 2340 2400 2460 2520 2580 ggaccagtac ctgtaccage tgcgcaccca tcacctgage caaatcactg aggetgeect 2640 ggccctgaag ctcggccaca gtgagctccc tgcagctctg gagcaggcgg aggactggct 2700 cetgegtetg egtgeeetgg cagaggagee ecagaacage etgeeggaca tegteatetg 2760 gatgetgeag ggagacaage gtgtggeata ccagegggtg ccegeceace aagteetett 2820 ctcccggcgg ggtgccaact actgtggcaa gaattgtggg aagctacaga caatctttct 2880 gaaatatccg atggagaagg tgcctggcgc ccggatgcca gtgcagatac gggtcaagct gtggtttggg ctctctgtgg atgagaagga gttcaaccag tttgctgagg ggaagctgtc 2940 3000 tgtetttget gaaacetatg agaacgagae taagttggee ettgttggga actggggeae 3060 aacgggcctc acctaccca agttttctga cgtcacgggc aagatcaagc tacccaagga 3120 cagcitccgc ccctcggccg gctggacctg ggctggagat tggttcgtgt gtccggagaa gactctgctc catgacatgg acgccggtca cctgagcttc gtggaagagg tgtttgagaa 3180 3240 ccagaccegg cttecceggag gecagtggat ctacatgagt gacaactaca cegatgtgaa 3300 cggggagaag gtgcttccca aggatgacat tgagtgccca ctgggctgga agtgggaaga tgaggaatgg tccacagacc tcaaccggc tgtcgatgag caaggctggg agtatagcat caccatccc ccggagcga agccgaagca ctgggtccct gctgagaaga tgtactacac 3360 3420 3480 acaccgacgg cggcgctggg tgcgcctgcg caggagggat ctcagccaaa tggaagcact 3540 gaaaaggcac aggcaggcgg aggcggaggg cgagggctgg gagtacgcct ctctttttgg 3600 ctggaagtte cacetegagt accgcaagac agatgcette egeegeegee getggegeeg 3660 tegeatggag ceaetggaga agaeggggee tgeagetgtg tttgeeettg agggggeeet gggeggegtg atggatgaca agagtgaaga tteeatgtee gteteeacet tgagettegg 3720 3780 tgtgaacaga cccacgattt cctgcatatt cgactatggg aaccgctacc atctacgctg 3840 ctacatgtac caggeceggg acctggetge gatggacaag gactetttt ctgateeta 3900 tgccatcgtc tccttcctgc accagagcca gaagacggtg gtggtgaaga acacccttaa 3960 cccacctgg gaccagacgc tcatcttcta cgagatcgag atctttggcg agccggccac agttgctgag caaccgcca gcattgtggt ggagctgtac gaccatgaca cttatggtgc 4020 4080 agacgagttt atgggtcgct gcatctgtca accgagtctg gaacggatgc cacggctggc 4140 ctggttccca ctgacgaggg gcagccagcc gtcgggggag ctgctggcct cttttgagct 4200 catccagaga gagaagccgg ccatccacca tattcctggt tttgaggtgc aggagacatc 4260 aaggatectg gatgagtetg aggacacaga ectgeectae ecaceacece agagggagge 4320 caacatctac atggttcctc agaacatcaa gccagcgctc cagcgtaccg ccatcgagat cctggcatgg ggcctgcgga acatgaagag ttaccagctg gccaacatct cctcccccag 4380 4440 cetegtggta gagtgtgggg gccagacggt gcagtectgt gtcatcagga acetecggaa 4500 4560 gaaccccaac tttgacatet gcaccetett catggaagtg atgetgeeca gggaggaget ctactgccc cccatcaccg tcaaggtcat cgataaccgc cagtttggcc gccggcctgt ggtgggccag tgtaccatcc gctccctgga gagcttcctg tgtgacccct actcggcgga gagtccatcc ccacagggtg gcccagacga tgtgagccta ctcagtcctg gggaagacgt 4620 4680 4740 getcategae attgatgaca aggageceet catececate caggaggaag agtteatega 4800 ttggtggagc aaattetttg cetecatagg ggagagggaa aagtgegget cetacetgga gaaggatttt gacacectga aggtetatga cacacagetg gagaatgtgg aggeetttga 4860 4920 gggcctgtct gacttttgta acaccttcaa gctgtaccgg ggcaagacgc aggaggagac agaagatca tetgtgattg gtgaatttaa gggcctcttc aaaatttatc ccctcccaga agacccagcc atccccatgc ccccaagaca gttccaccag ctggccgccc agggacccca 4980 5040 5100 ggagtgettg gteegtatet acattgteeg ageatttgge etgeageeca aggaceceaa 5160 tggaaagtgt gatccttaca tcaagatctc catagggaag aaatcagtga gtgaccagga 5220 taactacate ceetgeacge tggageeegt atttggaaag atgttegage tgacetgeac 5280 tetgeetetg gagaaggace taaagateae tetetatgae tatgacetee tetecaagga egaaaagate ggtgagaegg tegtegaeet ggagaacagg etgetgteea agtttgggge 5340 5400 tegetgtgga etcecacaga cetactgtgt etctggaceg aaccagtgge gggaccaget 5460

```
ccgccctcc cagctcctcc acctcttctg ccagcagcat agagtcaagg cacctgtgta
                                                                                       5520
ccggacagac cgtgtaatgt ttcaggataa agaatattcc attgaagaga tagaggctgg
                                                                                       5580
caggatecca aacceacace tgggeecagt ggaggagegt etggetetge atgtgettea
                                                                                       5640
gcagcaggge ctggtcccgg agcacgtgga gtcacggccc ctctacagcc ccctgcagcc agacatcgag caggggaagc tgcagatgtg ggtcgaccta tttccgaagg ccctggggcg gcctggacct cccttcaaca tcaccccacg gagagccaga aggtttttcc tgcgttgtat
                                                                                       5700
                                                                                       5760
                                                                                       5820
                                                                                       5880
tatctggaat accagagatg tgatcctgga tgacctgagc ctcacggggg agaagatgag
cgacatttat gtgaaaggtt ggatgattgg ctttgaagaa cacaagcaaa agacagacgt
                                                                                       5940
gcattatcgt tccctgggag gtgaaggcaa cttcaactgg aggttcattt tccccttcga ctacctgcca gctgagcaag tctgtaccat tgccaagaag gatgccttct ggaggctgga caagactgag agcaaaatcc cagcacgagt ggtgttccag atctgggaca atgacaagtt
                                                                                       6000
                                                                                       6060
                                                                                       6120
ctcctttgat gattttctgg gctccctgca gctcgatctc aaccgcatgc ccaagccagc
                                                                                       6180
                                                                                       6240
caagacagec aagaagtget cettggacca getggatgat getttecaec cagaatggtt
                                                                                       6300
tgtgtccctt tttgagcaga aaacagtgaa gggctggtgg ccctgtgtag cagaagaggg
tgagaagaaa atactggcgg gcaagctgga aatgaccttg gagattgtag cagagagtga gcatgaggag cggcctgctg gccagggccg agatgagccc aacatgaacc ctaagcttga
                                                                                       6360
                                                                                       6420
ggacccaagg cgccccgaca cctccttcct gtggtttacc tccccataca agaccatgaa
                                                                                       6480
                                                                                       6540
gttcatcctg tggcggcgtt tccggtgggc catcatcctc ttcatcatcc tcttcatcct
gctgctgttc ctggccatct tcatctacgc cttcccgaac tatgctgcca tgaagctggt
                                                                                       6600
gaagecette agetgaggae teteetgeee tgtagaaggg geegtggggt eeeetceage
                                                                                       6660
atgggactgg cctgcctcct ccgcccagct cggcgagctc ctccagacct cctaggcctg
                                                                                       6720
attgtcctgc cagggtgggc agacagacag atggaccggc ccacactccc agagttgcta
                                                                                       6780
acatggaget etgagateae eccaetteca teattteett etcecceaae ecaacgettt
                                                                                       6840
tttggatcag ctcagacata tttcagtata aaacagttgg aaccacaaaa aaaaaaaaa
                                                                                       6900
                                                                                       6911
aaaaaaaaa a
```

<210> 18 <211> 6911 <212> DNA <213> Homo sapiens

<400> 18

60 tegacegece agecaggtge aaaatgeegt gteattggga gaeteegeag eeggageatt 120 agattacago togacogago togggaagog oggoggogt ggaagatgag cagaagocco 180 geceactgga geageeggg gtggecegtt cecetttaag ageaactget etaageeagg agecagagat tegageegge etegeeage eageeetete eagegaggg acceacaage 240 300 ggegeetegg ceeteeegae ettteegage eetetttgeg eeetgggege aeggggeeet 360 420 acacgogoca agcatgotga gggtottoat cototatgoo gagaacgtoo acacaccoga 480 caccgacate agegatgeet actgeteege ggtgtttgca ggggtgaaga agagaaccaa agtcatcaag aacagcgtga accetgtatg gaatgaggga titigaatggg accetcaaggg 540 catcccctg gaccagggct ctgagcttca tgtggtggtc aaagaccatg agacgatggg 600 gaggaacagg ttcctggggg aagccaaggt cccactccga gaggtcctcg ccacccctag 660 720 tetgteegee agetteaatg ecceetget ggacaccaag aageageeea caggggeete gctggtcctg caggtgtcct acacaccgct gcctggagct gtgcccctgt tcccgccccc tactcctctg gagccctccc cgactctgcc tgacctggat gtagtggcag acacaggagg agaggaagac acagaggacc agggactcac tggagatgag gcggagccat tcctggatca 780 840 900 aagcggagge cegggggete ceaceacec aaggaaacta cetteacgte etecgeecea 960 ctacccggg atcaaaagaa agcgaagtgc gcctacatct agaaagctgc tgtcagacaa 1020 accgcaggat ttccagatca gggtccaggt gatcgagggg cgccagctgc cgggggtgaa 1080 1140 catcaagect gtggtcaagg ttaccgctgc agggcagacc aageggacge ggatccacaa gggaaacage ccactettea atgagaetet tttetteaac ttgtttgaet eteetgggga 1200 gctgtttgat gagcccatct ttatcacggt ggtagactct cgttctctca ggacagatgc 1260 tetecteggg gagtteegga tggaegtggg caccatttae agagageece ggeaegeeta teteaggaag tggetgetge teteagaeee tgatgaette tetgetgggg ecagaggeta 1320 1380 cctgaaaaca agcctitgig tgctggggcc tggggacgaa gcgcctcigg agagaaaaga 1440 ccctctgaa gacaaggagg acattgaaag caacctgctc cggcccacag gcgtagccct 1500 1560 gcgaggagcc cacttctgcc tgaaggtctt ccgggccgag gacttgccgc agatggacga tgccgtgatg gacaacgtga aacagatett tggettegag agtaacaaga agaacttggt 1620 ggacccetti gtggaggtca gettigeggg gaaaatgetg tgeageaaga tettggagaa gaeggeeaac eeteagtgga accagaacat cacactgeet geeatgtte eetecatgtg 1680 1740 1800 cgaaaaaatg aggattegta teatagactg ggacegeetg acteacaatg acategtgge taccacctac ctgagtatgt cgaaaatctc tgcccctgga ggagaaatag aagaggagcc 1860 tgcaggtgct gtcaagcctt cgaaagcctc agacttggat gactacctgg gcttcctccc cacttttggg ccctgctaca tcaacctcta tggcagtccc agagagttca caggcttccc 1920 1980 agaccectae acagagetea acacaggeaa gggggaaggt gtggettate gtggeegget 2040 tetgetetee etggagacca agetggtgga geacagtgaa cagaaggtgg aggacettee 2100

2160

tgcggatgac atcetecggg tggagaagta cettaggagg egeaagtact eeetgtttge ggeettetae teagceacea tgetgeagga tgtggatgat geeatceagt ttgaggteag 2220 catcgggaac tacgggaaca agttcgacat gacctgcctg ccgctggcct ccaccactca gtacagccgt gcagtctttg acgggtgcca ctactactac ctaccctggg gtaacgtgaa 2280 2340 acetgiggig gigetgicai ceiaciggga ggacateage catagaaicg agacteagaa 2400 ccagctgctt gggattgctg accggctgga agctggcctg gagcaggtcc acctggccct 2460 gaaggcgcag tgctccacgg aggacgtgga ctcgctggtg gctcagctga cggatgagct catcgcaggc tgcagccagc ctctgggtga catccatgag acaccctctg ccaccacct 2520 2580 ggaccagtac etgtaccage tgegcaccca teacetgage caaateactg aggetgeest ggccetgaag eteggcaca gtgagetece tgeagetetg gageaggegg aggactgget 2640 2700 cctgcgtctg cgtgccctgg cagaggagcc ccagaacagc ctgccggaca tcgtcatctg 2760 gatgctgcag ggagacaagc gtgtggcata ccagcgggtg cccgcccacc aagtcctctt 2820 ctcccggcgg ggtgccaact actgtggcaa gaattgtggg aagctacaga caatctttct gaaatatccg atggagaagg tgcctggcgc ccggatgcca gtgcagatac gggtcaagct 2880 2940 gtggtttggg ctctctgtgg atgagaagga gttcaaccag tttgctgagg ggaagctgtc 3000 tgtctttgct gaaacctatg agaacgagac taagttggcc cttgttggga actggggcac 3060 aacgggcctc acctaccca agttttctga cgtcacgggc aagatcaagc tacccaagga 3120 cagetteege eceteggeeg getggaeetg ggetggagat tggttegtgt gteeggagaa 3180 gactetgete catgacatgg acgeeggtea cetgagette gtggaagagg tgtttgagaa 3240 ccagacccgg cttcccggag gccagtggat ctacatgagt gacaactaca ccgatgtgaa 3300 ccagaccegg cttccca aggatgacat tgagtgccca ctgggctgga agtgggaaga tgaggaatgg tccacagacc tcaaccggc tgtcggtgccct caggatgacat tgcgccatccc ccggagcga agccgaagca ctggggtccct ggtggaagaa tgtactacac acaccgacgg tgcgcctggg cgcgctggg ccaggagggat ctcagccaaa tggaagcact ccagcacac caggagggat ctcagccaca cfctttttcc 3360 3420 3480 3540 gaaaaggcac aggcaggcgg aggcggaggg cgagggctgg gagtacgcct ctctttttgg ctggaagttc cacctcgagt accgcaagac agatgccttc cgccgccgcc gctggcgccg 3600 3660 tegeatggag ceaetggaga agaeggggee tgeagetgtg titgeeettg agggggeeet gggeggegtg atggatgaca agagtgaaga tteeatgtee gteteeaeet tgagettegg 3720 3780 tgtgaacaga cccacgattt cctgcatatt cgactatggg aaccgctacc atctacgctg ctacatgtac caggcccggg acctggctgc gatggacaag gactctttt ctgatcccta 3840 3900 tgccatcgtc tccttcctgc accagagcca gaagacggtg gtggtgaaga acacccttaa ccccacctgg gaccagacgc tcatcttcta cgagatcgag atctttggcg agccggccac 3960 4020 agttgctgag caaccgccca gcattgtggt ggagctgtac gaccatgaca cttatggtgc 4080 agacgagttt atgggtcgct gcatctgtca accgagtctg gaacggatgc cacggctggc 4140 ctggttccca ctgacgaggg gcagccagcc gtcgggggag ctgctggcct cttttgagct catccagaga gagaagccgg ccatccacca tattcctggt tttgaggtgc aggagacatc 4200 4260 aagggteetg gatgagtetg aggacacaga cetgeeetae ecaceacece agagggagge caacatetae atggtteete agaacateaa gecagegete eagegtaceg ceategagat 4320 4380 cctggcatgg ggcctgcgga acatgaagag ttaccagctg gccaacatet cetccccag 4440 cctcgtggta gagtgtgggg gccagacggt gcagtcctgt gtcatcagga acctccggaa 4500 gaaccccaac tttgacatct gcacctctt catggaagtg atgctgccca gggaggagct ctactgccc cccatcaccg tcaaggtcat cgataaccgc cagtttggcc gccggcctgt ggtgggccag tgtaccatcc gctcctgga gagcttcctg tgtgacccct actcggcgga 4560 4620 4680 gagtecatec ccacagggtg geccagaega tgtgageeta etcagteetg gggaagaegt getcategae attgatgaea aggageecet cateccate caggaggaag agtteatega 4740 4800 ttggtggagc aaattctttg cctccatagg ggagagggaa aagtgcggct cctacctgga 4860 4920 gaaggatttt gacaccctga aggtctatga cacacagctg gagaatgtgg aggcctttga gggcctgtct gacttttgta acaccttcaa gctgtaccgg ggcaagacgc aggaggagac agaagatcca tctgtgattg gtgaatttaa gggcctcttc aaaatttatc ccctcccaga 4980 5040 agacccagec atecccatge ceccaagaca gitecaccag etggeegeec agggacceca 5100 ggagtgettg gteegtatet acattgteeg ageatttgge etgeageeca aggaceecaa 5160 tggaaagtgt gatcettaca tcaagatete catagggaag aaatcagtga gtgaccagga taactacate ceetgeacge tggagecegt atttggaaag atgttegage tgacctgeac tetgeetetg gagaaggace taaagateae tetetatgae tatgacetee tetecaagga eggagaaggae tegtegaeet ggagaacagg etgetgteea agtttgggge 5220 5280 5340 5400 tegetgtgga eteccacaga cetactgtgt etetggaceg aaccagtgge gggaceaget 5460 cegecectee cagetectee acetettetg ceageageat agagteaagg cacetgtgta 5520 ccggacagac cgtgtaatgt ttcaggataa agaatattcc attgaagaga tagaggctgg 5580 caggatecca aacccacace tgggeeccagt ggaggagegt etggetetge atgtgettea 5640 gcagcaggge ctggtcccgg agcacgtgga gtcacggccc ctctacagcc ccctgcagcc agacatcgag caggggaagc tgcagatgtg ggtcgaccta tttccgaagg ccctggggcg 5700 5760 gcctggacct cccttcaaca tcaccccacg gagagccaga aggtttttcc tgcgttgtat 5820 tatctggaat accagagatg tgatcctgga tgacctgagc ctcacggggg agaagatgag cgacatttat gtgaaaggtt ggatgattgg ctttgaagaa cacaagcaaa agacagacgt 5880 5940 gcattatcgt tccctgggag gtgaaggcaa cttcaactgg aggttcattt tcccttcgg ctacctgcca gctgagcaag tctgtaccat tgccaagaag gatgccttct ggaggctgga caagactgag agcaaaatcc cagcacgagt ggtgttccag atctgggaca atgacaagtt 6000 6060 6120

```
ctcctttgat gattttctgg gctccctgca gctcgatctc aaccgcatgc ccaagccagc caagacagcc aagaagtgct ccttggacca gctggatgat gctttccacc cagaatggtt
                                                                                                  6180
                                                                                                  6240
tgtgtccctt tttgagcaga aaacagtgaa gggctggtgg ccctgtgtag cagaagaggg
                                                                                                  6300
tgagaagaaa atactggcgg gcaagctgga aatgaccttg gagattgtag cagagagtga gcatgaggag cggcctgctg gccagggccg ggatgagccc aacatgaacc ctaagcttga ggacccaagg cgccccgaca cctccttcct gtggtttacc tccccataca agaccatgaa
                                                                                                  6360
                                                                                                  6420
                                                                                                  6480
                                                                                                  6540
gttcatcctg tggcggcgtt tccggtgggc catcatcctc ttcatcatcc tcttcatcct
                                                                                                  6600
getgetgtte etggecatet teatetacge ettecegaae tatgetgeea tgaagetggt
gaagecette agetgaggae teteetgeee tgtagaaggg geegtggggt eeeeteeage atgggaetgg cetgeeteet eegeeeaget eggegagete etceagaeet eetaggeetg
                                                                                                  6660
                                                                                                  6720
attgtcctgc cagggtgggc agacagacag atggaccggc ccacactccc agagttgcta
                                                                                                  6780
acatggaget etgagateae eccaetteea teattteett eteccecaae ccaacgettt
                                                                                                  6840
tttggatcag ctcagacata tttcagtata aaacagttgg aaccacaaaa aaaaaaaaa
                                                                                                  6900
                                                                                                  6911
aaaaaaaaa a
```

<210> 19 <211> 6911 <212> DNA <213> Homo sapiens

<400> 19

tcgaccgccc agccaggtgc aaaatgccgt gtcattggga gactccgcag ccggagcatt 60 120 agattacage tegacggage tegggaaggg eggegggggt ggaagatgag cagaageece tgttctcgga acgccggctg acaagcgggg tgagcgcagg cggggggggg acccagccta gcccactgga gcagccggg gtggcccgtt cccctttaag agcaactgct ctaagccagg agccagagat tcgagccggc ctcgcccagc cagccctctc cagcgagggg acccacaagc 180 240 300 ggcgcctcgg ccctcccgac ctttccgagc cctctttgcg ccctgggcgc acggggccct 360 acacgogoca agoatgotga gggtottoat cototatgoo gagaacgtoo acacaccoga 420 caccgacatc agegatgect actgeteege ggtgtttgea ggggtgaaga agagaaccaa agteateaag aacagegtga accetgtatg gaatgaggga tttgaatggg accetcaaggg catececetg gaccaggget etgagettea tgtggtggte aaagaccatg agacgatggg 480 540 600 gaggaacagg ttcctggggg aagccaaggt cccactccga gaggtcctcg ccaccctag 660 tetgteegee agetteaatg ecceetget ggacaccaag aageageea caggggeete getggteetg caggtgteet acacaceget geetggaget gtgeeeetgt teeegeeee tacteetetg gageeeteee egactetgee tgacetggat gtagtggeag acacaggagg 720 780 840 agaggaagac acagaggacc agggactcac tggagatgag gcggagccat tcctggatca 900 aageggagge cegggggete ceaceacec aaggaaacta cetteaegte eteegeecea 960 ctacccggg atcaaagaa agcgaagtgc gcctacatct agaaagctgc tgtcagacaa 1020 accgcaggat ttccagatca gggtccaggt gatcgagggg cgccagctgc cgggggtgaa 1080 1140 catcaagect gtggtcaagg ttaccgetge agggeagace aageggaege ggatceacaa gggaaacage ceactettea atgagactet titetteaac tigtitgaet etectgggga 1200 getgtttgat gageceatet ttateaeggt ggtagaetet egttetetea ggacagatge 1260 tctcctcggg gagttccgga tggacgtggg caccatttac agagagcccc ggcacgccta 1320 tctcaggaag tggctgctgc tctcagaccc tgatgacttc tctgctgggg ccagaggcta 1380 cctgaaaaca agcctttgtg tgctggggcc tggggacgaa gcgcctctgg agagaaaaga 1440 cccctctgaa gacaaggagg acattgaaag caacctgctc cggcccacag gcgtagccct gcgaggagcc cacttctgcc tgaaggtctt ccgggccgag gacttgccgc agatggacga 1500 1560 tgccgtgatg gacaacgtga aacagatctt tggcttcgag agtaacaaga agaacttggt 1620 ggacccetti gtggaggtca gettigeggg gaaaatgetg tgeageaaga tettggagaa gaeggeeaae eeteagtgga accagaacat cacactgeet geeatgtte eetecatgtg 1680 1740 cgaaaaaatg aggattegta teatagactg ggacegeetg acteacaatg acategtgge 1800 1860 taccacctac ctgagtatgt cgaaaatctc tgcccctgga ggagaaatag aagaggagcc tgcaggtgct gtcaagcctt cgaaagcctc agacttggat gactacctgg gcttcctccc 1920 cacttttggg ccctgctaca tcaacctcta tggcagtccc agagagttca caggcttccc 1980 agacccctac acagagetea acacaggeaa gggggaaggt gtggettate gtggeegget 2040 tetgetetee etggagacea agetggtgga geacagtgaa cagaaggtgg aggacettee 2100 tgcggatgac atcctccggg tggagaagta ccttaggagg cgcaagtact ccctgtttgc 2160 ggccttctac tcagccacca tgctgcagga tgtggatgat gccatccagt ttgaggtcag 2220 2280 categggaac taegggaaca agttegacat gaeetgeetg eegetggeet eeaceactea gtacagccgt gcagtctttg acgggtgcca ctactactac ctaccctggg gtaacgtgaa 2340 acctgtggtg gtgctgtcat cctactggga ggacatcagc catagaatcg agactcagaa ccagctgctt gggattgctg accggctgga agctggcctg gagcaggtcc acctggccct 2400 2460 gaaggcgcag tgctccacgg aggacgtgga ctcgctggtg gctcagctga cggatgagct 2520 categoagge tgcagecage etetgggtga catecatgag acacectetg ecacecacet 2580 ggaccagtac ctgtaccage tgegeaceca teacetgage caaateactg aggetgeeet 2640 ggccctgaag ctcggccaca gtgagctccc tgcagctctg gagcaggcgg aggactggct 2700 cetgegtetg egtgeeetgg cagaggagee ceagaacage etgeeggaea tegteatetg 2760

gatgctgcag ggagacaagc gtgtggcata ccagcgggtg cccgcccacc aagtcctctt 2820 ctcccggcgg ggtgccaact actgtggcaa gaattgtggg aagctacaga caatctttct 2880 gaaatateeg atggagaagg tgeetggege ceggatgeea gtgeagatae gggteaaget 2940 gtggtttggg ctctctgtgg atgagaagga gttcaaccag tttgctgagg ggaagctgtc tgtctttgct gaaacctatg agaacgagac taagttggcc cttgttggga actggggcac 3000 3060 aacgggcctc acctaccca agttttctga cgtcacgggc aagatcaagc tacccaagga 3120 cagetteege ceeteggeeg getggaeetg ggetggagat tggttegtgt gteeggagaa gaetetgete catgaeatgg acgeeggtea eetgagette gtggaagagg tgtttgagaa 3180 3240 ccagacccgg cttcccggag gccagtggat ctacatgagt gacaactaca ccgatgtgaa cgggggagaag gtgcttccca aggatgacat tgagtgccca ctgggctgga agtgggaaga 3300 3360 tgaggaatgg tecacagace tcaaceggge tgtegatgag caaggetggg agtatageat 3420 caccatecee eeggagegga ageegaagea etgggteeet getgagaaga tgtactaeae 3480 acaccgacgg cggcgctggg tgcgcctgcg caggagggat ctcagccaaa tggaagcact gaaaaggcac aggcaggcgg aggcggaggg cgagggctgg gagtacgcct ctctttttgg ctggaagttc cacctcgagt accgcaagac agatgccttc cgccgccgcc gctggcgccg 3540 3600 3660 tegeatggag ceaetggaga agaeggggee tgeagetgtg tttgeeettg agggggeeet 3720 gggcggcgtg atggatgaca agagtgaaga ttccatgtcc gtctccacct tgagcttcgg 3780 tgtgaacaga cccacgattt cctgcatatt cgactatggg aaccgctacc atctacgctg 3840 ctacatgtac caggeceggg acetggetge gatggacaag gaetetttt etgateeta tgecategte teetteetge aceagageca gaagaeggtg gtggtgaaga acaceettaa ececacetgg gaecagaege teatetteta egagategag atetttggeg ageeggeeae 3900 3960 4020 agttgctgag caaccgccca gcattgtggt ggagctgtac gaccatgaca cttatggtgc 4080 agacgagttt atgggteget geatetgtea accgagtetg gaacggatge cacggetgge 4140 ctggttccca ctgacgaggg gcagccagcc gtcgggggag ctgctggcct cttttgagct catccagaga gagaagccgg ccatccacca tattcctggt tttgaggtgc aggagacatc 4200 4260 aaggateetg gatgagtetg aggacacaga cetgeeetae ecaceacee agagggagge caacatetae atggtteete agaacateaa gecagegete cagegtaceg ceategagat 4320 4380 cetggcatgg ggcetgegga acatgaagag ttaccagetg gccaacatet cetececcag cetegtggta gagtgtgggg gccagacggt gcagtcetgt gtcatcagga aceteeggaa gaaceccaac tttgacatet gcacectett catggaagtg atgetgecca gggaggaget ctactgecce cecateaceg tcaaggtcat cagatacege cagtttggec geeggeetgt 4440 4500 4560 4620 ggtgggccag tgtaccatcc gctccctgga gagcttcctg tgtgacccct actcggcgga 4680 gagtecatec ccacagggtg geccagacga tgtgagecta etcagtectg gggaagaegt 4740 getcategae attgatgaca aggageceet catececate caggaggaag agtteatega 4800 ttggtggagc aaattctttg cctccatagg ggagagggaa aagtgcggct cctacctgga gaaggatttt gacaccctga aggtctatga cacacagctg gagaatgtgg aggcctttga 4860 4920 gggcctgtct gacttttgta acaccttcaa gctgtaccgg ggcaagacgc aggaggagac agaagatcca tctgtgattg gtgaatttaa gggcctcttc aaaatttatc ccctcccaga 4980 5040 5100 agaccagcc atcccatgc ccccaagaca gttccaccag ctggccgccc agggacccca ggagtgettg gteegtatet acattgteeg ageatttgge etgeageeca aggaceecaa 5160 tggaaagtgt gatccttaca tcaagatctc catagggaag aaatcagtga gtgaccagga 5220 taactacate ceetgeacge tggagecegt atttggaaag atgttegage tgacetgeac 5280 taactacate ceetycaege tygageege attragalay atytetyage tyaectyaege tetgeetety gagaaggace taaagateae tetetatgae tatgacetee tetecaagga eggagaacagg tegtegaeet eggagaacagg etgetyteea ectetyty etgetyteea ectetyty engageege teggagaege tetetgaeeg eaceagtyge gygaecaget ecggeeetee eggeteetee ectetyty ecageagae eggagaege agagteaagg eacetytyta eggagaegae eggagaegae eggagaege eggagaege engageege engagaege eggagaege engagaege 5340 5400 5460 5520 5580 5640 caggatecca aacccacace tgggcccagt ggaggagegt etggetetge atgtgettea gcagcaggge ctggtcccgg agcacgtgga gtcacggccc ctctacagcc ccctgcagcc agacatcgag caggggaagc tgcagatgtg ggtcgaccta tttccgaagg ccctggggcg gcctggacct cccttcaaca tcaccccacg gagagccaga aggtttttcc tgcgttgtat tatctggaat accagagatg tgatcctgga tgacctgagc ctcacggggg agaagatgag 5700 5760 5820 5880 cgacatttat gtgaaaggtt ggatgattgg ctttgaagaa cacaagcaaa agacagacgt 5940 gcattatcgt tccctgggag gtgaaggcaa cttcaactgg aggttcattt tccccttcga 6000 ctacetgeca getgageaag tetgtaceat tgecaagaag gatgeettet ggaggetgga 6060 caagactgag agcaaaatcc cagcacgagt ggtgttccag atctgggaca atgacaagtt 6120 ctcctttgat gattttctgg gctccctgca gctcgatctc aaccgcatgc ccaagccagc caagacagcc aagaagtgct ccttggacca gctggatgat gctttccacc cagaatggtt 6180 6240 tgtgtccctt tttgagcaga aaacagtgaa gggctggtgg ccctgtgtag cagaagaggg tgagaagaaa atactggcgg gcaagctgga aatgaccttg gagattgtag cagaagagtga 6300 6360 gcatgaggag cggcctgctg gccagggccg ggatgagccc aacatgaacc ctaagcttga ggacccaagg cgccccgaca cctccttcct gtggtttacc tccccataca agaccatgaa 6420 6480 gttcatcctg tggcggtgtt tccggtgggc catcatcctc ttcatcatcc tcttcatcct 6540 getgetgtte etggecatet teatetacge ettecegaae tatgetgeca tgaagetggt 6600 gaagecette agetgaggae teteetgeee tgtagaaggg geegtggggt eeeeteeage atgggaetgg cetgeeteet eegeceaget eggegagete etceagacet ectaggeetg attgteetge eagggtgge agacagacag atggaeegge ceacacteee agagttgeta 6660 6720 6780

```
acatggaget etgagateae eccaetteea teattteett etecceeaae ecaaegettt
                                                                             6840
6900
                                                                             6911
aaaaaaaaa a
        <210> 20
       <211> 6911
        <212> DNA
        <213> Homo sapiens
        <400> 20
                                                                               60
togaccqccc agccaggtgc aaaatgccgt gtcattggga gactccgcag ccggagcatt
                                                                              120
agattacage tegaeggage tegggaaggg eggegggggt ggaagatgag cagaageece
                                                                               180
týttotogga acgoeggotg acaagegggg tgagegeagg eggggegggg acceagecta
gcccactgga gcagccggg gtggcccgtt cccctttaag agcaactgct ctaagccagg agccagagat tcgagccggc ctcgcccagc cagccctctc cagcgagggg acccacaagc
                                                                              240
                                                                               300
ggcgcctcgg ccctcccgac ctttccgagc cctctttgcg ccctgggcgc acggggccct
                                                                              360
                                                                              420
acacgegeca ageatgetga gggtetteat cetetatgee gagaacgtee acacacega
caccaacate agegatgeet actgeteege ggtgtttgea ggggtgaaga agagaaccaa
                                                                              480
agtcatcaag aacagcgtga accetgtatg gaatgaggga tttgaatggg accetaaggg cateccectg gaccaggget ctgagettea tgtggtggte aaagaccatg agacgatggg gaggaacagg tteetggggg aagccaaggt eccaeteega gaggteeteg ecaeceetag
                                                                              540
                                                                              600
                                                                              660
                                                                               720
totatocace agetteaata ecceetact agacaceaag aageageeca cagagacete
                                                                              780
getggteetg caggtgteet acacecget geetggaget gtgeecetgt teeegeeeee
tactcctctg gagccctccc cgactctgcc tgacctggat gtagtggcag acacaggagg agaggaagac acagaggacc agggactcac tggagatgag gcggagccat tcctggatca
                                                                              840
                                                                              900
aagcggagge cegggggete ceaceacece aaggaaacta cetteaegte eteegececa
                                                                              960
ctacccggg atcaaaagaa agcgaagtgc gcctacatct agaaagctgc tgtcagacaa
                                                                             1020
                                                                             1080
accgcaggat ttccagatca gggtccaggt gatcgagggg cgccagctgc cgggggtgaa
                                                                             1140
catcaagect gtggtcaagg ttaccgetge agggcagace aageggacge ggatecacaa
                                                                             1200
gggaaacage ceactettea atgagaetet titetteaac tigtitgaet etectgggga
gctgtttgat gagcccatct ttatcacggt ggtagactct cgttctctca ggacagatgc
                                                                             1260
tetecteggg gagtteegga tggaegtggg caccatttae agagageece ggeaegeeta
                                                                             1320
                                                                             1380
tetcaggaag tggetgetge tetcagacce tgatgactte tetgetgggg ccagaggeta
                                                                             1440
cctgaaaaca agcctttgtg tgctggggcc tggggacgaa gcgcctctgg agagaaaaga
cccctctgaa gacaaggagg acattgaaag caacctgctc cggcccacag gcgtagccct gcgaggagcc cacttctgcc tgaaggtctt ccgggccgag gacttgccgc agatggacga
                                                                             1500
                                                                             1560
                                                                             1620
tgccgtgatg gacaacgtga aacagatctt tggcttcgag agtaacaaga agaacttggt
ggacccttt gtggaggtca gctttgcggg gaaaatgctg tgcagcaaga tcttggagaa
                                                                             1680
gacggccaac cctcagtgga accagaacat cacactgcct gccatgtttc cctccatgtg
                                                                             1740
                                                                             1800
cgaaaaaatg aggattcgta tcatagactg ggaccgcctg actcacaatg acatcgtggc
taccacctac ctgagtatgt cgaaaatctc tgcccctgga ggagaaatag aagaggagcc
                                                                             1860
tgcaggtgct gtcaagcctt cgaaagcctc agacttggat gactacctgg gcttcctccc cacttttggg ccctgctaca tcaacctcta tggcagtccc agagagttca caggcttccc
                                                                             1920
                                                                             1980
agacccctac acagagetea acacaggeaa gggggaaggt gtggettate gtggeegget
                                                                             2040
                                                                             2100
totgototoc otggagacca agotggtgga gcacagtgaa cagaaggtgg aggacottoc
tgcggatgac atcctccggg tggagaagta ccttaggagg cgcaagtact ccctgtttgc
                                                                             2160
ggccttctac tcagccacca tgctgcagga tgtggatgat gccatccagt ttgaggtcag
                                                                              2220
                                                                              2280
categggaac taegggaaca agttegacat gacetgeetg eegetggeet ecaceactea
gtacagccgt gcagtctttg acgggtgcca ctactactac ctaccctggg gtaacgtgaa
                                                                             2340
acctgtggtg gtgctgtcat cctactggga ggacatcagc catagaatcg agactcagaa
                                                                             2400
ccagctgctt gggattgctg accggctgga agctggcctg gagcaggtcc acctggccct
                                                                              2460
                                                                              2520
gaaggcgcag tgctccacgg aggacgtgga ctcgctggtg gctcagctga cggatgagct
                                                                              2580
categeagge tgeageeage etetgggtga cateeatgag acaeectetg ceacceacet
ggaccagtac ctgtaccage tgegeaccea teacetgage caaateactg aggetgeeet
                                                                              2640
ggccctgaag ctcggccaca gtgagctccc tgcagctctg gagcaggcgg aggactggct
                                                                              2700
cetgegtetg egtgeeetgg cagaggagee ecagaacage etgeeggaca tegteatetg
                                                                              2760
gatgctgcag ggagacaagc gtgtggcata ccagcgggtg cccgcccacc aagtcctctt
                                                                              2820
ctcccggcgg ggtgccaact actgtggcaa gaattgtggg aagctacaga caatctttct gaaatatccg atggagaagg tgcctggcgc ccggatgcca gtgcagatac gggtcaagct
                                                                              2880
                                                                              2940
gtggtttggg ctctctgtgg atgagaagga gttcaaccag tttgctgagg ggaagctgtc
                                                                              3000
tgtctttgct gaaacctatg agaacgagac taagttggcc cttgttggga actggggcac
                                                                              3060
                                                                              3120
aacqqqcctc acctacccca agttttctga cgtcacgggc aagatcaagc tacccaagga
cagetteege eecteggeeg getggaeetg ggetggagat tggttegtgt gteeggagaa
                                                                              3180
gactotgoto catgacatgg acgooggtoa cotgagotto gtggaagagg tgtttgagaa
                                                                              3240
ccagaccegg cttcccggag gccagtggat ctacatgagt gacaactaca ccgatgtgaa
                                                                              3300
cggggagaag gtgcttccca aggatgacat tgagtgccca ctgggctgga agtgggaaga
                                                                              3360
                                                                              3420
tgaggaatgg tccacagacc tcaaccgggc tgtcgatgag caaggctggg agtatagcat
```

```
caccatcccc ccggagcgga agccgaagca ctgggtccct gctgagaaga tgtactacac
                                                                                                  3480
acaccgacgg cggcgctggg tgcgcctgcg caggagggat ctcagccaaa tggaagcact
                                                                                                  3540
gaaaaggcac aggcaggcgg aggcggaggg cgagggctgg gagtacgcct ctctttttgg ctggaagttc cacctcgagt accgcaagac agatgccttc cgccgccgcc gctggcgccg
                                                                                                  3600
                                                                                                  3660
tegeatggag ceaetggaga agaeggggee tgeagetgtg tttgeeettg agggggeeet
                                                                                                  3720
gggcggcgtg atggatgaca agagtgaaga ttccatgtcc gtctccacct tgagcttcgg
                                                                                                  3780
tgtgaacaga cccacgattt cctgcatatt cgactatggg aaccgctacc atctacgctg
                                                                                                  3840
ctacatgtac caggeceggg acctggetge gatggacaag gactetttt ctgateeta
                                                                                                  3900
tgccatcgtc tccttcctgc accagagcca gaagacggtg gtggtgaaga acacccttaa ccccacctgg gaccagacgc tcatcttcta cgagatcgag atctttggcg agccggccac
                                                                                                  3960
                                                                                                  4020
agttgctgag caaccgccca gcattgtggt ggagctgtac gaccatgaca cttatggtgc
                                                                                                  4080
agacgagttt atgggteget geatetgtea accgagtetg gaacggatge cacggetgge
                                                                                                  4140
ctggttccca ctgacgaggg gcagccagcc gtcgggggag ctgctggcct cttttgagct catccagaga gagaagccgg ccatccacca tattcctggt tttgaggtgc aggagacatc
                                                                                                   4200
                                                                                                  4260
aaggateetg gatgagtetg aggacacaga cetgeeetae ecaceacee agagggagge caacatetae atggtteete agaacateaa gecagegete cagegtaeeg ceategagat
                                                                                                   4320
                                                                                                   4380
cctggcatgg ggcctgcgga acatgaagag ttaccagctg gccaacatct cctcccccag cctcgtggta gagtgtgggg gccagacggt gcagtcctgt gtcatcagga acctccggaa gaaccccaac tttgacatct gcacctctt catggaagtg atgctgcca gggaggagct
                                                                                                   4440
                                                                                                   4500
                                                                                                   4560
ctactgccc cccatcaccg tcaaggtcat cgataaccgc cagtttggcc gccggcctgt ggtgggccag tgtaccatcc gctccctgga gagettcctg tgtgacccct actcggcgga
                                                                                                   4620
                                                                                                   4680
gagtecatec ccacagggtg geccagaega tgtgageeta etcagteetg gggaagaegt getcategae attgatgaea aggageecet catececate caggaggaag agtteatega
                                                                                                   4740
                                                                                                   4800
ttggtggagc aaattetttg cetecatagg ggagagggaa aagtgegget cetacetgga
                                                                                                   4860
gaaggatttt gacaccctga aggtctatga cacacagctg gagaatgtgg aggcctttga
                                                                                                   4920
gggcetgtet gaettttgta acacetteaa getgtacegg ggcaagaege aggaggagae
                                                                                                   4980
 agaagateca tetgtgattg gtgaatttaa gggeetette aaaatttate eeeteecaga
                                                                                                   5040
 agacccagcc atccccatgc ccccaagaca gttccaccag ctggccgccc agggacccca
                                                                                                   5100
 ggagtgettg gteegtatet acattgteeg ageatttgge etgeageeca aggaceecaa tggaaagtgt gateettaca teaagatete catagggaag aaateagtga gtgaceagga
                                                                                                   5160
                                                                                                   5220
 taactacate ecetgeacge tggagecegt atttggaaag atgttegage tgacetgeac
                                                                                                   5280
tetgeetetg gagaaggace taaagateae tetetatgae tatgacetee tetecaagga egaaaagate ggtgagaegg tegtegacet ggagaacagg etgetgteea agtttgggge tegetgtgga eteccacaga ectaetgtgt etetggaceg aaceagtgge gggaceaget
                                                                                                   5340
                                                                                                   5400
                                                                                                   5460
 ccgcccctcc cagetectcc acetettetg ccageageat agagteaagg cacetgtgta
                                                                                                   5520
 ccggacagac cgtgtaatgt ttcaggataa agaatattcc attgaagaga tagaggctgg
                                                                                                   5580
 caggatecca aacccacace tgggcccagt ggaggagegt etggetetge atgtgettea
                                                                                                   5640
 gcagcaggge etggtecegg agcacgtgga gtcacggcce etctacagee ecetgcagee
                                                                                                   5700
 agacatcgag caggggaagc tgcagatgtg ggtcgaccta tttccgaagg ccctggggcg gcctggacct cccttcaaca tcaccccacg gagagccaga aggtttttcc tgcgttgtat
                                                                                                   5760
                                                                                                    5820
 tatotggaat accagagatg tgatootgga tgacotgago otcacggggg agaagatgag
                                                                                                   5880
 cgacatttat gtgaaaggtt ggatgattgg ctttgaagaa cacaagcaaa agacagacgt
                                                                                                    5940
 gcgttatcgt tccctgggag gtgaaggcaa cttcaactgg aggttcattt tccccttcga ctacctgcca gctgagcaag tctgtaccat tgccaagaag gatgccttct ggaggctgga
                                                                                                    6000
                                                                                                    6060
 caagactgag agcaaaatcc cagcacgagt ggtgttccag atctgggaca atgacaagtt
                                                                                                    6120
 ctcctttgat gattttctgg gctccctgca gctcgatctc aaccgcatgc ccaagccagc caagacagcc aagaagtgct ccttggacca gctggatgat gctttccacc cagaatggtt
                                                                                                    6180
                                                                                                    6240
 tgtgtccctt tttgagcaga aaacagtgaa gggctggtgg ccctgtgtag cagaagaggg
tgagaagaaa atactggcgg gcaagctgga aatgaccttg gagattgtag cagaagatga
gcatgaggag cggcctgctg gccagggccg ggatgagccc aacatgaacc ctaagcttga
                                                                                                    6300
                                                                                                    6360
                                                                                                    6420
 ggacccaagg cgcccgaca cctccttcct gtggtttacc tccccataca agaccatgaa
                                                                                                    6480
 gttcatectg tggeggegtt teeggtggge cateatecte tteateatec tetteatect getgetgtte etggecatet teatetacge ettecegaae tatgetgeea tgaagetggt
                                                                                                    6540
                                                                                                    6600
 gaageeette agetgaggae teteetgeee tgtagaaggg geegtggggt eeeeteeage atgggaetgg cetgeeteet eegeceaget eggegagete etceagaeet eetaggeetg
                                                                                                    6660
                                                                                                    6720
 attgteetge cagggtggge agacagacag atggacegge ecacaetece agagttgeta
                                                                                                    6780
  acatggaget etgagateae eccaetteca teattteett eteccecaae ecaaegettt
                                                                                                    6840
 tttggatcag ctcagacata tttcagtata aaacagttgg aaccacaaaa aaaaaaaaa
                                                                                                    6900
                                                                                                    6911
  aaaaaaaaa a
```

<210> 21

<211> 6909

<212> DNA

<213> Homo sapiens

<400> 21 tcgaccgccc agccaggtgc aaaatgccgt gtcattggga gactccgcag ccggagcatt

agattacage tegaeggage tegggaaggg eggegggggt ggaagatgag eagaageece 120 180 tgttctcgga acgccggctg acaagcgggg tgagcgcagg cggggggggg acccagccta geccaetgga geageeggg gtggeeegtt cecetttaag ageaactget etaageeagg agecagagat tegageegge etegeeeage cageeetete cagegagggg acceacaage 240 300 ggcgcctcgg ccctcccgac ctttccgagc cctctttgcg ccctgggcgc acggggccct 360 420 acacqcqcca agcatgctga gggtcttcat cctctatgcc gagaacgtcc acacacccga 480 caccgacatc agcgatgect actgeteege ggtgtttgca ggggtgaaga agagaaccaa agtcatcaag aacagcgtga accetgtatg gaatgaggga titgaatggg accetaaggg catcccctg gaccaggget ctgagettea tgtggtggte aaagaccatg agacgatggg gaggaacagg tecetggggg aagccaaggt cccactccga gaggtceteg ccacccctag 540 600 660 720 tetigteegee agetteaatg ecceeetget ggacaccaag aageageeca caggggeete 780 getggteetg caggtgteet acacaceget geetggaget gtgeecetgt teeegeeece tactcctctg gagccctccc cgactctgcc tgacctggat gtagtggcag acacaggagg agaggaagac acagaggacc agggactcac tggagatgag gcggagccat tcctggatca 840 900 aageggagge eegggggete ceaceacec aaggaaacta cetteacgte etecgececa 960 ctacccggg atcaaaagaa agcgaagtgc gcctacatct agaaagctgc tgtcagacaa 1020 1080 accgcaggat ttccagatca gggtccaggt gatcgagggg cgccagctgc cgggggtgaa catcaageet gtggteaagg ttacegetge agggeagace aageggaege ggatecacaa 1140 gggaaacage ceactettea atgagactet titetteaac tigtitgaet etectgggga 1200 gctgtttgat gagcccatct ttatcacggt ggtagactct cgttctctca ggacagatgc 1260 tctcctcggg gagttccgga tggacgtggg caccatttac agagagcccc ggcacgccta 1320 tetcaggaag tggetgetge tetcagacce tgatgactte tetgetgggg ceagaggeta 1380 1440 cctgaaaaca agcctttgtg tgctggggcc tggggacgaa gcgcctctgg agagaaaaga ccctctgaa gacaaggagg acattgaaag caacctgctc cggcccacag gcgtagccct gcgaggagcc cacttctgcc tgaaggtctt ccgggccgag gacttgccgc agatggacga 1500 1560 tgccgtgatg gacaacgtga aacagatett tggcttegag agtaacaaga agaacttggt 1620 ggacccettt gtggaggtca gctttgcggg gaaaatgctg tgcagcaaga tcttggagaa gacggccaac cctcagtgga accagaacat cacactgcct gccatgtttc cctccatgtg 1680 1740 cgaaaaaatg aggattegta teatagactg ggacegeetg acteacaatg acategtgge taccacetae etgagtatgt egaaaatete tgeeeetgga ggagaaatag aagaggagee 1800 1860 tgcaggtgct gtcaagcctt cgaaagcctc agacttggat gactacctgg gcttcctccc 1920 1980 cacttttggg coctgctaca tcaacctcta tggcagtccc agagagttca caggcttccc agacccctac acagagetca acacaggeaa gggggaaggt gtggettate gtggeegget tetgetetee etggagacca agetggtgga geacagtgaa cagaaggtgg aggacettee tgeggatgac atecteeggg tggagaagta cettaggagg egcaagtact ecetgtttge ggeettetae teageeacca tgetgeagga tgtggatgat geeateeagt ttgaggteag 2040 2100 2160 2220 catcgggaac tacgggaaca agttcgacat gacctgcctg ccgctggcct ccaccactca 2280 2340 gtacagccgt gcagtctttg acgggtgcca ctactactac ctaccctggg gtaacgtgaa acctgtggtg gtgctgtcat cctactggga ggacatcagc catagaatcg agactcagaa ccagctgctt gggattgctg accggctgga agctggcctg gagcaggtcc acctggccct gaaggcgcag tgctccacgg aggacgtgga ctcgctggtg gctcagctga cggatgagct 2400 2460 2520 2580 categeagge tgcagecage etetgggtga categatgag acacectetg ccacecacet ggaccagtac ctgtaccage tgcgcaccca tcacctgage caaatcactg aggetgeeet 2640 ggccctgaag ctcggccaca gtgagctccc tgcagctctg gagcaggcgg aggactggct cctgcgtctg cgtgccctgg cagaggagcc ccagaacagc ctgccggaca tcgtcatctg gatgctgcag ggagacaagc gtgtggcata ccagcgggtg cccgcccacc aagtcctctt 2700 2760 2820 ctcccggcgg ggtgccaact actgtggcaa gaattgtggg aagctacaga caatctttct 2880 gaaatatccg atggagaagg tgcctggcgc ccggatgcca gtgcagatac gggtcaagct 2940 gtggtttggg ctctctgtgg atgagaagga gttcaaccag tttgctgagg ggaagctgtc tgtctttgct gaaacctatg agaacgagac taagttggcc cttgttggga actggggcac aacgggcctc acctaccca agttttctga cgtcacgggc aagatcaagc tacccaagga 3000 3060 3120 3180 caqcttccqc ccctcqqccq gctggacctg ggctggagat tggttcgtgt gtccggagaa gactetgete catgacatgg acgccggtca cctgagette gtggaagagg tgtttgagaa 3240 ccagaccegg cttcccggag gccagtggat ctacatgagt gacaactaca ccgatgtgaa cggggagaag gtgcttccca aggatgacat tgagtgccca ctgggctgga agtgggaaga tgaggaatgg tccacagacc tcaaccgggc tgtcgatgag caaggctggg agtatagcat 3300 3360 3420 caccatecce eeggagegga ageegaagea etgggteeet getgagaaga tgtactacac acacegaegg eggegetggg tgegeetgeg eaggagggat eteageeaaa tggaageaet 3480 3540 gaaaaggcac aggcaggcgg aggcggaggg cgagggctgg gagtacgcct ctctttttgg ctggaagttc cacctcgagt accgcaagac agatgccttc cgccgccgcc gctggcgccg 3600 3660 tcgcatggag ccactggaga agacggggcc tgcagctgtg tttgcccttg agggggccct 3720 gggcggcgtg atggatgaca agagtgaaga ttccatgtcc gtctccacct tgagcttcgg 3780 tgtgaacaga cccacgattt cctgcatatt cgactatggg aaccgctacc atctacgctg 3840 ctacatgtac caggcccggg acctggctgc gatggacaag gactcttttt ctgatccctatgccatcgtc tccttcctgc accagagcca gaagacggtg gtggtgaaga acacccttaaccccacctgg gaccagacgc tcatcttcta cgagatcgag atctttggcg agccggccac 3900 3960 4020 agttgctgag caaccgccca gcattgtggt ggagctgtac gaccatgaca cttatggtgc 4080

```
agacgagttt atgggtcgct gcatctgtca accgagtctg gaacggatgc cacggctggc
                                                                                     4140
ctggttccca ctgacgaggg gcagccagcc gtcgggggag ctgctggcct cttttgagct catccagaga gagaagccgg ccatccacca tattcctggt tttgaggtgc aggagacatc
                                                                                     4200
                                                                                     4260
aaggateetg gatgagtetg aggacacaga cetgeeetae ceaceacee agagggagge caacatetae atggtteete agaacateaa geeagegete eagegtaeeg ceategagat
                                                                                     4320
                                                                                     4380
cctggcatgg ggcctgcgga acatgaagag ttaccagctg gccaacatct cctccccag
                                                                                     4440
cctcgtggta gagtgtgggg gccagacggt gcagtcctgt gtcatcagga acctccggaa
                                                                                    4500
gaaccccaac tttgacatct gcaccctctt catggaagtg atgctgccca gggaggagct
                                                                                    4560
ctactgccc cccatcaccg tcaaggtcat cgataaccgc cagtttggcc gccggcctgt ggtgggccag tgtaccatcc gctccctgga gagcttcctg tgtgacccct actcggcgga
                                                                                     4620
                                                                                    4680
                                                                                     4740
gagtecatec ccacagggtg geccagaega tgtgageeta etcagteetg gggaagaegt
gctcatcgac attgatgaca aggagcccct catccccatc caggaggaag agttcatcga ttggtggagc aaattctttg cctccatagg ggagagggaa aagtgcggct cctacctgga gaaggatttt gacaccctga aggtctatga cacacagctg gagaatgtgg aggcctttga
                                                                                    4800
                                                                                     4860
                                                                                     4920
gggcctgtct gacttttgta acaccttcaa gctgtaccgg ggcaagacgc aggaggagac
                                                                                     4980
                                                                                     5040
agaagatcca totgtgattg gtgaatttaa gggcototto aaaatttato coctoccaga
agacccagec atccccatge ceccaagaca gttecaccag etggeegeec agggacceca
                                                                                    5100
ggagtgcttg gtccgtatct acattgtccg agcatttggc ctgcagccca aggaccccaa
                                                                                     5160
tggaaagtgt gatccttaca tcaagatctc catagggaag aaatcagtga gtgaccagga
                                                                                     5220
taactacate ecctgeacge tggagecegt atttggaaag atgttegage tgacetgeac
                                                                                     5280
tetgeetetg gagaaggace taaagateae tetetatgae tatgacetee tetecaagga
                                                                                     5340
                                                                                    5400
cgaaaagate qqtqaqaegg tegtegaeet ggagaacagg etgetgteea agtttgggge
tegetgtgga eteceacaga cetactgtgt etetggaceg aaccagtgge gggaccaget
                                                                                    5460
cegecetee cageteetee acetettetg ceageageat agagteaagg cacetgtgta
                                                                                     5520
ccggacagac cgtgtaatgt ttcaggataa agaatattcc attgaagaga tagaggctgg
                                                                                     5580
caggatecca aacceacace tgggeecagt ggaggagegt etggetetge atgtgettea
                                                                                     5640
gcagcagggc ctggtcccgg agcacgtgga gtcacggccc ctctacagcc ccctgcagcc
                                                                                     5700
agacategag caggggaage tgcagatgtg ggtcgaceta tttccgaagg ccctggggcg
                                                                                     5760
gcctggacct cccttcaaca tcaccccacg gagagccaga aggtttttcc tgcgttgtat tatctggaat accagagatg tgatcctgga tgacctgagc ctcacggggg agaagatgag
                                                                                     5820
                                                                                     5880
                                                                                     5940
cgacatttat gtgaaaggtt ggatgattgg ctttgaagaa cacaagcaaa agacagacgt
gcattatcgt tccctgggag gtgaaggcaa cttcaactgg aggttcattt tccccttcga ctacctgca gctgagcaag tctgtaccat tgccaagaag gatgcettct ggaaggctgga caagactgag caaaatcca gcacgagtgg tgttccagat ctgggacaat gacaagttct cctttgatga ttttctgggc tccctgcagc tcgatctaa ccgcatgccc aagccagcca
                                                                                     6000
                                                                                     6060
                                                                                     6120
                                                                                     6180
agacagecaa gaagtgetee ttggaccage tggatgatge tttecaccea gaatggtttg
                                                                                     6240
tgtccctttt tgagcagaaa acagtgaagg gctggtggcc ctgtgtagca gaagagggtg
                                                                                     6300
agaagaaaat actggcgggc aagctggaaa tgaccttgga gattgtagca gagagtgagc
                                                                                     6360
atgaggageg geetgetgge cagggeeggg atgageecaa catgaaceet aagettgagg
                                                                                     6420
acccaaggcg coccgacace tectteetgt ggtttacete eccatacaag accatgaagt
                                                                                     6480
teatectgtg geggegttte eggtgggeea teatectett cateatecte tteatectge
                                                                                     6540
                                                                                     6600
tgctgttcct ggccatcttc atctacgcct tcccgaacta tgctgccatg aagctggtga
agecetteag etgaggaete teetgeeetg tagaagggge egtggggtee eetecageat
                                                                                     6660
gggactggcc tgcctcctcc gcccagctcg gcgagctcct ccagacctcc taggcctgat
                                                                                     6720
tgtcctgcca gggtgggcag acagacagat ggaccggccc acactcccag agttgctaac
                                                                                     6780
                                                                                     6840
atggagetet gagateacce cacttecate attteettet eccecaacce aacgettttt
tggatcagct cagacatatt tcagtataaa acagttggaa ccacaaaaaa aaaaaaaaa
                                                                                     6900
                                                                                     6909
aaaaaaaa
         <210> 22
         <211> 20
         <212> DNA
         <213> Homo sapiens
         <400> 22
                                                                                        20
 tgggacctca agggcatccc
         <210> 23
         <211> 20
         <212> DNA
         <213> Homo sapiens
         <400> 23
                                                                                         20
  accatgctgc aggatgtgga
```

BNSDOCID: <WO\_\_\_\_\_0011157A1\_I\_>

<210> 24 <211> 20

	2> DNA 3> Homo	ganie	ens				
400> gggaggtga	)> 24	cttca					20
gggaggcga	. ggcaa	CCCCa					
	> 25						
	> 20						
	2> DNA 3> Homo	ganie	na				
~21.	, HOO	Bapie					
	> 25	_					20
ctcacgggg	g agaag	atgag					20
<21	> 26						
<21	L> 20						
	2> DNA						
<21.	3> Homo	sapıe	ens				
<40	> 26						
ctgtggcgg	gtttc	cggtg					20
	> 27						
	L> 20 2> DNA						
	3> Homo	sapie	ens				
		•					
	)> 27						20
acatcaagg	a tcctg	gatga					20
<21	)> 28						
	1> 20						
	2> DNA						
<21	3> Homo	sapie	ens				
<40	)> 28						
ctgtggcgg	gitte	cggtg					20
.04							
	0> 29 1> 20						
	2> DNA						
	3> Homo	sapie	ens				
		•					
	0> 29						20
acagacgtg	c attat	cgttc					20
<21	0> 30						
	1> 20						
	2> DNA	_					
<21	3> Homo	sapie	ens				
<40	0> 30						
aagactgag	a gcaaa	atccc					20
-01	0. 21						
	0> 31 1> 507						
	2> DNA						
	3> Homo	sapi	ens				
	0> 31	aataa	aaaatgccgt	atcattage?	dacticedese	ccagagcatt	60
agattagag	c tease	agaac	tcgggaaggg	cggcagaaat	ggaagatgag	cagaaqcccc	120
tattctcaa	a acqco	egacta	acaagcgggg	tgagcgcagg	cggggcgggg	acccagccta	180
gcccactgg	a gcago	caaaa	gtggcccgtt	cccctttaag	agcaactgct	ctaagccagg	240
agccagaga	t tcgac	accaac	ctcqcccaqc	cagccctctc	cagcgagggg	acccacaagc	300
ggcgcctcg	g ccctc	ccgac	ctttccgagc gggtcttcat	cctctatccc	dadaacdtcc	acacacccaa	360 420
acacycycc	u aycal	-yya	999000000	Joseph	J-2		

```
caccgacatc agegatgeet actgeteege ggtgtttgea ggtaggaggg geegaeeaee
                                                                          480
                                                                          507
ctcgccgggg tcggggtggg gtagagg
      <210> 32
      <211> 183
      <212> DNA
      <213> Homo sapiens
      <400> 32
aaaggeggga tgtgtetete catteteeet tttgtgtete ttgtaggggt gaagaagaga
                                                                           60
accaaagtca tcaagaacag cgtgaaccct gtatggaatg aggtatgtga gtttttctcc
                                                                          120
                                                                          180
ttoottitot ototgtotgo tgcagggggo ttgggaggag gtgccttoto agcagtgtoo
                                                                           183
      <210> 33
      <211> 264
      <212> DNA
      <213> Homo sapiens
      <400> 33
                                                                           60
cattcatgaa tgcctactca gtgccctggt ggcacgaagg tgaaccagac acagtctctt
ctcctagagg gccataggtt aagatgcctt ttctcttttt cttccaggga tttgaatggg
                                                                           120
acctcaaggg catecectg gaccaggget etgagettea tgtggtggte aaagaccatg agacgatggg gaggaacagg taaggtggee agaggggggt getecatgge ttgaaggtge
                                                                           180
                                                                           240
                                                                           264
aggtaggatt gtggagtata caga
       <210> 34
       <211> 223
       <212> DNA
       <213> Homo sapiens
cagaagagee agggtgeett aggetagttt tetacatttg aettetetet eeteteaggt
       <400> 34
                                                                            60
                                                                           120
tectggggga agecaaggte ceaeteegag aggteetege caeeectagt etgteegeca
getteaatge ecceetgetg gacaccaaga ageageeeac aggggtaagt geceateage
                                                                           180
                                                                           223
ctctgccagg ttaaggtcca aggcattgcc aggtggcttc ctc
       <210> 35
       <211> 224
       <212> DNA
       <213> Homo sapiens
       <400> 35
cagtggtccg aggccagcgc accaacctgt cccccacgtc tcatctcttc caggcctcgc
                                                                            60
tggtcctgca ggtgtcctac acaccgctgc ctggagctgt gcccctgttc ccgcccccta
                                                                           120
ctectetgga geeeteeceg actetgeetg acctggatgt agtggeaggt gggtageeca
                                                                           180
                                                                           224
cgttggcctg gctgggcccc agcaagaatg gccggcagtg gcac
       <210> 36
       <211> 315
       <212> DNA
       <213> Homo sapiens
       <400> 36
 aggggcaggg gcagggccag agggccaggc ctcattaggg ccctctcctc ttagacacag
                                                                            60
 gaggagagga agacacagag gaccagggac tcactggaga tgaggcggag ccattcctgg
                                                                           120
 atcaaagegg aggeeegggg geteecacea ecceaaggaa actacettea egteeteege
                                                                           180
 cccactacce cgggatcaaa agaaagegaa gtgegeetae atetagaaag etgetgteag
                                                                           240
 acaaaccgca ggatttccag gtgatgaacg ggctttctct gaccccaggc tcctcttcag
                                                                           300
                                                                           315
 ccatcagctg cgggt
       <210> 37
       <211> 249
```

<213> Homo sapiens

<212> DNA

<pre>&lt;400&gt; 37 ccagtggtga gatggtccct actcttcccc cttctggctt ggggtgaaca tcaagcctgt atccacaagg gaaacagccc accttgtgg</pre>	tcagatcagg ggtcaaggtt	gtccaggtga accgctgcag	tcgaggggcg ggcagaccaa	ccagctgccg gcggacgcgg	60 120 180 240 249
<210> 38 <211> 185 <212> DNA <213> Homo sapie	ns				
<400> 38 ccctggcctg agggatcagc cttttcttca acttgtttga gtatgtctca gcagtcaaag cacac	ctctcctaga	gagetgtttg	atgagcccat	ctttatcacg	60 120 180 185
<210> 39 <211> 196 <212> DNA <213> Homo sapie	ns				
<pre>&lt;400&gt; 39 aagagctatt gggttggccg ctctgcaggt ggtagactct taattgctta ttttctaaaa ggaccatggg cagggg</pre>	cgttctctca	ggacagatgc	tctcctcggg	gagttccggg	60 120 180 196
<210> 40 <211> 178 <212> DNA <213> Homo sapie	ens				
<400> 40 tggaatcgta taatgcacca tctcctctct tgattgcaga accactttgg ccgtatcctt	tggacgtggg	caccatttac	agagagcccc	gtgagttctc	60 120 178
<210> 41 <211> 231 <212> DNA <213> Homo sapie	ens				
<pre>&lt;400&gt; 41 ggggtcttct gattctggga aggaagtggc tgctgctctc aaaacaagcc tttgtgtgct gtcttcctta cggtccccca</pre>	agaccctgat ggggcctggg	gacttctctg gacgaagcgc	ctggggccag ctgtgagtac	aggetacetg attteeetgg	60 120 180 231
<210> 42 <211> 247 <212> DNA <213> Homo sapie	ens				
<400> 42 gtcaaaaccc tgtgctcagg gagaaaagac ccctctgaag cgtagccctg cgaggagccc gagtgcgtgg ggcgcccct tgggggt	acaaggagga acttctgcct	gaaggtcttc	cgggccgagg	ggcccacagg acttgccgca	60 120 180 240 247
<210> 43 <211> 179					

<211> 179 <212> DNA <213> Homo sapiens

<400> 43 caggcagtga ctggtgtgtc cagatctttg gcttcgagag tttgcgggga aaatggtaag	taacaagaag	aacttqqtgg	acccctttgt	ggaggtcagc	60 120 179
<210> 44 <211> 202 <212> DNA <213> Homo sapi	ens				
<400> 44 ccccggggga gcccagagtc agctgtgcag caagatcttg tgcctgccat ggtgagcctc cttcgggagg tccagggctc	gagaagacgg ctgtccccag	ccaaccctca	gtggaaccag	dacatcacac	60 120 180 202
<210> 45 <211> 167 <212> DNA <213> Homo sapi	ens				
<pre>&lt;400&gt; 45 gggaggggct gttctatctt agtttccctc catgtgcgaa ttggagtctt tagggcgggc</pre>	aaaatqagga	ttcgtatcat	agactggtga	tteetteete gttetgagte	60 120 167
<210> 46 <211> 220 <212> DNA <213> Homo sapi	ens				
<pre>&lt;400&gt; 46 tgtggcctga gttcctttcc cctgactcac aatgacatcg tggaggagaa atagaaggta cccccctcta tccagcttac</pre>	tggctaccac tgttccctct	tegttetgee	atyttyaaaa	ccccgcccc	60 120 180 220
<210> 47 <211> 172 <212> DNA <213> Homo sapi	.ens				
<pre>&lt;400&gt; 47 acagcctgtt catgtaaccc ttctttacgc ttcagaggag cgttgctgtc accttggggag</pre>	, cctacaaata	ctgtcaagec	Licyanaycc	ccagactgta	60 120 172
<210> 48 <211> 200 <212> DNA <213> Homo sapi	ens				
<pre>&lt;400&gt; 48 ccgacccctc tgattgccac cttttgggcc ctgctacatc acccctacac agagctcaac atgccacatg cccaggtggg</pre>	aacctctatg acaggcaagg	gcagtcccag	agagttcaca	ggetteeeag	60 120 180 200
<210> 49 <211> 217 <212> DNA <213> Homo sap	Lens				
<pre>&lt;400&gt; 49 cctcccctct gtctcccctc cttatcgtgg ccggcttctc aggtggagga ccttcctgc</pre>	r ctctccctaa	agaccaagct	ggtggagcac	agugaadaga	60 120 180

gtgggagetg ggegtegggg caggg	aaggg atggcca			217
<210> 50 <211> 269 <212> DNA <213> Homo sapiens				
<pre>&lt;400&gt; 50 agcctgggtg cctttctttg ctcct gtaccttagg aggcgcaagt actcc ggatgtggat gatgccatcc agttt catgacctgc ctgccgctgg cctcc ggcagtgctc ctggctggga ccccg</pre>	ctgtt tgcggccttc gaggt cagcatcggg accac tcagtacagc	tactcagcca d	ccatgctgca acaagttcga ttgacggtga	60 120 180 240 269
<210> 51 <211> 225 <212> DNA <213> Homo sapiens				
<pre>&lt;400&gt; 51 actcctggca cagcgctcag gcccgg cctggggtaa cgtgaaacct gtggtc gaatcgagac tcagaaccag ctgctg ccaaagctgc acatgcctat gcatg</pre>	ggtgc tgtcatccta tggga ttgctgaccg	ctgggaggac a gctggtgagt	atcagccata gaaaacttgc	60 120 180 225
<210> 52 <211> 227 <212> DNA <213> Homo sapiens				
<pre>&lt;400&gt; 52 gggtccagca tgcaccctct gccctc caggaagctg gcctggagca ggtcc gtggactcgc tggtggctca gctgac cctggcgccc ctggtgccca cctctc</pre>	acctg gccctgaagg cggat gagctcatcg	cgcagtgctc c	cacggaggac gtagggggga	60 120 180 227
<210> 53 <211> 303 <212> DNA <213> Homo sapiens				
<pre>&lt;400&gt; 53 tgggagaccc tgggctcatc aggcgg ggtgacatcc atgagccaaat cactgg acccatcacc tgagccaaat cactgg ctccctgcag ctctggagca ggcgg gaggtaatta agcctggggg tgcct aac</pre>	ccacc cacctggacc aggct gccctggccc aggac tggctcctgc	agtacctgta of tgaagctcgg of gtctgcgtgc	ccagetgege ccacagtgag cctggcagag ggaacatcag	60 120 180 240 300 303
<210> 54 <211> 272 <212> DNA <213> Homo sapiens				
<pre>&lt;400&gt; 54 cgtgggcctg gtgtgtcacc atccc agcctgccgg acatcgtcat ctgga gtgcccgcc accaagtcct cttctc gggaagctac agacaatctt tctga tccaacataa ggcctttctc ccatc</pre>	tgctg cagggagaca cccgg cggggtgcca aagtg agttttcttt	agcgtgtggc a	ataccagcgg caagaattgt tgatcgtatt	60 120 180 240 272
<210> 55 <211> 219 <212> DNA <213> Homo sapiens				

<pre>&lt;400&gt; 55  tgtgggtttc tgtccttctt cggtacccag tatccgatgg agaaggtgcc tggcgcccgg atgccagtgc agatacgggt caagctgtgg tttgggctct ctgtggatga gaaggagttc aaccagtttg ctgaggggaa gctgtctgtc tttgctgaaa ccgtgagtac ctgccagccc ccacctctgc ctcccactac ctggagctgc cttggcccc</pre>	60 120 180 219
<210> 56 <211> 292 <212> DNA <213> Homo sapiens	
<pre>&lt;400&gt; 56  tgcctcccac tacctggagc tgccttggcc cccttcacgc ctcattcttc ctggccctcc agtatgagaa cgagactaag ttggcccttg ttgggaactg gggcacaacg ggcctcacct accccaagtt ttctgacgtc acgggcaaga tcaagctacc caaggacagc ttccgccct cggccggctg gacctgggct ggagattggt tcgtgtgtcc ggagaagacg tgagtcgtgg gcagggaggg ctggggagag ccaggccagg ctgcccacca tggactgcac cc</pre>	60 120 180 240 292
<210> 57 <211> 242 <212> DNA <213> Homo sapiens	
<pre>&lt;400&gt; 57 tggatgggg cctctccagc agagcagcag agactctgac cagccctcct ccacagtctg ctccatgaca tggacgccgg tcacctgagc ttcgtggaag aggtgtttga gaaccagacc cggcttcccg gaggccagtg gatctacatg agtgacaact acaccgatgt ggtaaagcag gcactcaggg gcaggtgggg tctagacatt tggtctctgg aggcacctgg tgctcaggga ca</pre>	60 120 180 240 242
<210> 58 <211> 215 <212> DNA <213> Homo sapiens	
<400> 58 tcacatctgt ctgtctcctc tcattgcttg cctgttcggt tttgtcctta gaacggggag aaggtgcttc ccaaggatga cattgagtgc ccactgggct ggaagtggga agatgaggaa tggtccacag acctcaaccg ggctgtcgat gagcaaggtg ggcagcatgt ggaacctggc gagccccatc cccggcaagc tctcaagcca tgcat	120
<210> 59 <211> 246 <212> DNA <213> Homo sapiens	
<400> 59 agagatggtc ccaggagaga tggggggaag tgccaagcaa tgagtgaccg gttcccctc ccccaggctg ggagtatagc atcaccatcc ccccggagcg gaagccgaag cactgggtcc ctgctgagaa gatgtactac acacaccgac ggcggcgctg ggtgcgcctg cgcaggaggg atctcagcca aatggaagca ctgaaaaaagg gtgagccagc aggtggtggg tgggagtgag gcctgt	180
<210> 60 <211> 253 <212> DNA <213> Homo sapiens	
<400> 60 cttcccaccg gcctctgagt ctgccccttc ttgtgcagca caggcaggcg gaggcggagg gcgagggctg ggagtacgcc tctctttttg gctggaagtt ccacctcgag taccgcaaga cagatgcctt ccgccgccgc cgctggcgcc gtcgcatgga gccactggag aagacggggccttgcagctgt gtttgccctt gagggggccc tggtatgtgg ggctgcactt gtcctggcttgggtagggta	180
<210> 61 <211> 177	

<212> DNA <213> Homo sapiens <400> 61 qaatctgcca taaccagctt cgtgtctcca gggcggcgtg atggatgaca agagtgaaga 60 120 ttccatqtcc qtctccacct tqagcttcgg tgtgaacaga cccacgattt cctgcatatt 177 cgactgtaag taggcttcga ggcctctatg gggtgataag ggtgtgtcac cttatgc <210> 62 <211> 181 <212> DNA <213> Homo sapiens <400> 62 aaccactcca gccactcact ctggcacctc tgttttttcc cttggtgaag atgggaaccg 60 120 ctaccatcta cgctgctaca tgtaccaggc ccgggacctg gctgcgatgg acaaggactc tttttctggt aggtgggaga gaggcaggag agtcagagac tgtgggctga gatctgggaa 180 181 <210> 63 <211> 319 <212> DNA <213> Homo sapiens <400> 63 ccccacatgg ctctggagaa gacatctctc agggtccctg ctgtgtaatg tctcccctcc 60 ccctctggcc atgcagatcc ctatgccatc gtctccttcc tgcaccagag ccagaagacg 120 gtggtggtga agaacacct taaccccacc tgggaccaga cgctcatctt ctacgagatc 180 gagatettig gegageegge cacagtiget gageaacege ceageatigt ggtggagetg tacgaceatg acacttatgt gagtetgeec ageteetgee tegteecete acagggaggg 240 300 319 accatgtgca aaggtgggg <210> 64 <211> 249 <212> DNA <213> Homo sapiens <400> 64 60 qccctqqqta agggatqctg attcttgtct ctctacgctt ggtctagggt gcagacgagt 120 ttatgggteg ctgcatctgt caaccgagte tggaacggat gecacggetg geetggttee cactgacgag gggcagccag ccgtcggggg agctgctggc ctcttttgag ctcatccaga 180 240 249 ggccaaccc <210> 65 <211> 158 <212> DNA <213> Homo sapiens <400> 65 60 cactgacata gtccatgagt gtcatgaggg tgatgggggc cttaggtgac aagcacatga ccagagetet etttettea etccageegg ccatecacea tatteetggt tttgaggtaa 120 gtottgetet gacettteet tetteaaact gattgeea 158 <210> 66 <211> 132 <212> DNA <213> Homo sapiens <400> 66 60 ctttttcccc ttccaaccc tctcaccatc tcctgatgtg cacatcccat ggctgtgggc 120 caggtgcagg agacatcaag gatcctggat gaggtgagct ggcggggccg aggtagaggg 132 aaggtgaagc ca <210> 67 <211> 216 <212> DNA

<213> Homo sapiens <400> 67 tetteettee acetttgtet ceattetace tgetgteeac tgeagtetga ggacacagae 60 ctgccctacc caccaccca gagggaggcc aacatctaca tggttcctca gaacatcaag 120 ccagcgctcc agcgtaccgc catcgaggtg agccgtccgg gcctgggcgt gggggctggg 180 216 agcageetge cetteccett cetggeecea geettt <210> 68 <211> 263 <212> DNA <213> Homo sapiens <400> 68 eccgggeett etgageeact etecteatte tgtgtgetta gaateetgge atggggeetg 60 120 cggaacatga agagttacca gctggccaac atctectece ccagectegt ggtagagtgt 180 gggggccaga cggtgcagtc ctgtgtcatc aggaacetec ggaagaacec caactttgac atetgeacce tetteatgga agtggtgage cecacetece tactgteece ttecagagte 240 263 ctggggctag aagttctaca tgt <210> 69 <211> 249 <212> DNA <213> Homo sapiens <400> 69 caggecagtg cgttcttcct cctccaccca gatgctgccc agggaggagc tctactgccc 60 cccatcacc gicaaggica tegataaccg ccagtitgge egeeggeetg tggtgggeea 120 gtgtaccate egeteetgg agagetteet gtgtgacee taeteggegg agagteeate 180 cccacagggt ggcccaggta ggggaagggg agatgatggg caggtcaggg aaggggggggc 240 249 ctagggcaa <210> 70 <211> 180 <212> DNA <213> Homo sapiens <400> 70 aggggcgagc cttttgagag agcccctgtc aggcctggat ggctccctcc cctgcagacg 60 atgtgagect actcagtect ggggaagacg tgeteatega cattgatgae aaggageece 120 tcatcccat ccaggtagga tgggcatcct ccagggaggc ctgggtcacc tttcccctcc 180 <210> 71 <211> 211 <212> DNA <213> Homo sapiens <400> 71 tgctgcttgg cgagtcctgt ttctgaaatg gtctctttct ttctacccac tcaggaggaa 60 gagttcatcg attggtggag caaattcttt gcctccatag gggagaggga aaagtgcggc 120 tectacetgg agaaggattt tgacacectg aaggtaagge etetetteag tetgacagte 180 211 ggtgtgtgtg tgcgtgctgg gcagtgggag a <210> 72 <211> 235 <212> DNA <213> Homo sapiens <400> 72 gttctacttt ctttctgtct cttgtcccct cctctaatcc ccatgtgtgg caggtctatg 60 acacacaget ggagaatgtg gaggeetttg agggeetgte tgaettttgt aacacettea 120 agetgtaceg gggcaagaeg caggaggaga cagaagatee atetgtgatt ggtgaattta 180 aggtaaatcc tcgaagacgt ccctaaccca ggtgggccta agactgtggt gttgg 235 <210> 73

<211> 268 <212> DNA

214

#### 41/68

#### <213> Homo sapiens <400> 73 ggggacacag ccaaaccata tcaacaatga tgataaaata aaattaaccc ttccttcttt tcagggcctc ttcaaaattt atcccctccc agaagaccca gccatcccca tgcccccaag 60 120 acagttccac cagetggccg cccagggacc ccaggagtgc ttggtccgta tctacattgt 180 ccgagcattt ggcctgcagc ccaaggaccc caatggaaag gtaactttct agagccctca 240 268 cctcccaga gtagcaggct caggtaca <210> 74 <211> 200 <212> DNA <213> Homo sapiens <400> 74 tttggaaagt gttttcacag aagtgttttg tctcctcctc cagtgtgatc cttacatcaa 60 gatctccata gggaagaaat cagtgagtga ccaggataac tacatcccct gcacgctgga 120 gcccgtattt ggaaagtaaa ttggggcatc ttgggtcttg gggtggagga gccagacagg 180 200 ataacccaca gtctagtggg <210> 75 <211> 263 <212> DNA <213> Homo sapiens <400> 75 cctgttccct tgggtgccct gtgttggctg acattcggga atctgcccct tcctgcagga 60 tgttcgagct gacctgcact ctgcctctgg agaaggacct aaagatcact ctctatgact 120 atgacetect etecaaggae gaaaagateg gtgagaeggt egtegaeetg gagaacagge 180 tgctgtccaa gtttggggct cgctgtggac tcccacagac ctactgtgtg tacgtggatg 240 ggggctggct gcctgcttct ctg 263 <210> 76 <211> 237 <212> DNA <213> Homo sapiens <400> 76 aagcateteg tetatgtett gtgettgete eteagetetg gaeegaacea gtggegggae 60 cageteegee ecteecaget ecteeacete ttetgeeage ageatagagt caaggeacet 120 gtgtaccgga cagaccgtgt aatgtttcag gataaagaat attccattga agagataggt 180 237 gagetgecae atgaceceaa accatggtgg getetegetg tatecetece tetetea <210> 77 <211> 245 <212> DNA <213> Homo sapiens <400> 77 tototogott coccaqctco tgcaactttt ttgtgttctc tctggggcag aggctggcag 60 120 gateceaaac ecacactgg geccagtgga ggagegtetg getetgeatg tgetteagea gragggretg gtcccggage acgtggagte acggcccete tacagecece tgcagecaga 180 catcgagcag gtaggacctt acccttggtc ccagagtcct cgaactccag aagcccaacc 240 245 ccagg <210> 78 <211> 214 <212> DNA <213> Homo sapiens ggtgcttggt aacagctggt taaatgagaa gggtggggag agaacggacc tgtctccgca 60 ggggaagetg gggaagetge agatgtgggt cgacctattt ccgaaggece tggggcggcc 120 tggacctccc ttcaacatca ccccacggag agccagaagg tgacttccca gccacaggct 180

ctgagctggg ctgaggggtg gggcgttgca gcct

<210> 79 <211> 229 <212> DNA <213> Homo sapien	18				
<pre>&lt;400&gt; 79 ttcttaaggc cttcccatcc t ccaggttttt cctgcgttgt a gcctcacggg ggagaagatg a tgcctgtcca gcttcccgca g</pre>	ittatctgga igcgacattt	ataccagaga atgtgaaagg	gtagggagcc	yatyattiya	60 120 180 229
<210> 80 <211> 261 <212> DNA <213> Homo sapien	ns				
<pre>&lt;400&gt; 80 acgatgtata tactgtgttg g tagttggatg attggctttg a gggaggtgaa ggcaacttca a gcaagtctgt accattgcca a cagggcttct aaagttagcc t</pre>	agaacacaa actggaggtt agaaggtcag	cattttcccc	ttcgactacc	tgccagctga	60 120 180 240 261
<210> 81 <211> 234 <212> DNA <213> Homo sapier	าธ				
<400> 81 tgcctctctc taactttgct t cttctggagg ctggacaaga o ggacaatgac aagttctcct t tagaatccca ttctgcacat o	ctgagagcaa Ltgatgattt	tctqqtqatt	ttctgggtaa	gcgctattgc	60 120 180 234
<210> 82 <211> 297 <212> DNA <213> Homo sapier	ns				
<400> 82 ggctacaggc tggcagtgat c cctgcagctc gatctcaacc c ggaccagctg gatgatgctt t agtgaagggc tggtggccct c tctacttcct ccagccccag	gcatgcccaa tccacccaga gtgtagcaga	atggtttgtg agagggtgag	tccctttttg aagaaaatac	agcagaaaac tggcggtaag	60 120 180 240 297
<210> 83 <211> 237 <212> DNA <213> Homo sapie	ns				
<400> 83 cctggttact ctccaggcca cccctcaggg caagctggaa ggcctgctgg ccagggccgg cagtgcccag ccctgagcc	atgaccttgg gatgagccca	agattgtagc acatgaaccc	agagagtgag taagcttgag	gacccaaggt	60 120 180 237
<210> 84 <211> 252 <212> DNA <213> Homo sapie	ns				
<pre>&lt;400&gt; 84 ccctagtaaa ggatgcccag ttcctgtggt ttacctcccc tgggccatca tcctctcat tacgccttcc cggtgagcag</pre>	atacaagacc catcctcttc	atgaagttca atcctgctgc	teetgtggeg tgtteetgge	catcttcatc	60 120 180 240

```
252
gggagttcat ca
      <210> 85
      <211> 391
      <212> DNA
      <213> Homo sapiens
      <400> 85
tggctgtgcc tgccccagtg ggatcaccat gggtccctgt ctcctccctc cctccagaac
                                                                           60
tatgctgcca tgaagctggt gaagcccttc agctgaggac tctcctgccc tgtagaaggg gccgtggggt cccctccagc atgggactgg cctgcctcct ccgcccagct cggcgagctc
                                                                          120
                                                                          180
ctccagacct cctaggcctg attgtcctgc cagggtgggc agacagacag atggaccggc
                                                                          240
ccacactccc agagttgcta acatggaget ctgagateac cccacttcca tcatttcctt
                                                                          300
                                                                          360
ctccccaac ccaacgcttt tttggatcag ctcagacata tttcagtata aaacagttgg
                                                                          391
aaccacaaaa aaaaaaaaaa aaaaaaaaaa a
      <210> 86
      <211> 51
      <212> PRT
      <213> Homo sapiens
      <400> 86
Lys Lys Arg Thr Lys Val Ile Lys Asn Ser Val Asn Pro Val Trp Asn
                                       10
Glu Gly Phe Glu Trp Asp Leu Lys Gly Ile Pro Leu Asp Gln Gly Ser
                                   25
              20
Glu Leu His Val Val Lys Asp His Glu Thr Met Gly Arg Asn Arg
         35
                               40
Phe Leu Gly
     50
      <210> 87
      <211> 45
      <212> PRT
      <213> Homo sapiens
      <400> 87
Ser Lys Ile Leu Glu Lys Thr Ala Asn Pro Gln Trp Asn Gln Asn Ile
                                        10
Thr Leu Pro Ala Met Phe Pro Ser Met Cys Glu Lys Met Arg Ile Arg
                                   25
              20
Ile Ile Asp Trp Asp Arg Leu Thr His Asn Asp Ile Val
          35
                               40
      <210> 88
      <211> 82
       <212> PRT
       <213> Homo sapiens
      <400> 88
Gln Ala Arg Asp Leu Ala Ala Met Asp Lys Asp Ser Phe Ser Asp Pro
                                        10
Tyr Ala Ile Val Ser Phe Leu His Gln Ser Gln Lys Thr Val Val Val
                                                         30
                                   25
              20
Lys Asn Thr Leu Asn Pro Thr Trp Asp Gln Thr Leu Ile Phe Tyr Glu
                               40
 Ile Glu Ile Phe Gly Glu Pro Ala Thr Val Ala Glu Gln Pro Pro Ser
                           55
                                                60
 Ile Val Val Glu Leu Tyr Asp His Asp Thr Tyr Gly Ala Asp Glu Phe
  65
Met Gly
       <210> 89
       <211> 79
```

BNSDOCID: <WO\_\_\_\_\_0011157A1\_I\_>

<212> PRT

<213> Homo sapiens

```
<400> 89
Ile Tyr Ile Val Arg Ala Phe Gly Leu Gln Pro Lys Asp Pro Asn Gly
                                                          15
                                      10
Lys Cys Asp Pro Tyr Ile Lys Ile Ser Ile Gly Lys Lys Ser Val Ser
                                  25
             20
Asp Gln Asp Asn Tyr Ile Pro Cys Thr Leu Glu Pro Val Phe Gly Lys
                              40
         35
Met Phe Glu Leu Thr Cys Thr Leu Pro Leu Glu Lys Asp Leu Lys Ile
                         55
Thr Leu Tyr Asp Tyr Asp Leu Leu Ser Lys Asp Glu Lys Ile Gly
 65
                      70
      <210> 90
      <211> 152
      <212> DNA
      <213> Homo sapiens
      <400> 90
acgatgtata tactgtgttg gaaatcttaa tgagaactat tctctaaaaa catgtatgtc
                                                                         60
tagttggatg attggctttg aagaacacaa gcaaaagaca gacgtgcatt atcgttccct
                                                                        120
                                                                        152
gggaggtgaa ggcaacttca actggaggtt ca
      <210> 91
      <211> 56
      <212> DNA
      <213> Homo sapiens
      <400> 91
gtcagtgtcc ttccgattcc ctgtggtgcc agcaccaggg cttctaaagt tagcct
                                                                         56
      <210> 92
      <211> 55
       <212> DNA
      <213> Homo sapiens
      <400> 92
tgcctctctc taactttgct tccttgcatc cttctctgtt cctcttccgg gtcag
                                                                         55
      <210> 93
      <211> 68
       <212> DNA
       <213> Homo sapiens
       <400> 93
gtaagcgcta ttgctagaat cccattctgc acatgggggc tgccccagaa cccacactgt
                                                                         60
                                                                          68
 gtgtttat
       <210> 94
       <211> 56
       <212> DNA
       <213> Homo sapiens
       <400> 94
ggctacaggc tggcagtgat cgagaaaccc ggccaaaaac cacctctctg ttgcag
                                                                          56
       <210> 95
       <211> 62
       <212> DNA
       <213> Homo sapiens
 gtaagtctac ttcctccagc cccagtggag ggcatggggg aagcttcttc catagaaatt
                                                                          60
                                                                          62
 gt
       <210> 96
       <211> 68
```

<212> DNA

	<213>	Homo	sapie	ens				
	<400>	96						60
cctggt		ctccag	ggcca	ctgagcagag	ccttcgtgcc	cctaaccaag	tgctctctgt	60 68
	<210>	97						
	<211>							
	<212>							
	<213>	Homo	sapie	ens				
	<400>						ananataan	59
gtcagt	gccc	agccc	ctgag	ccccaatgcc	cacaggcccg	ggggcacagg	cacagecca	•
	<210>	98						
	<211>							
	<212>							
	<213>	Homo	варт	ens				
	<400>							
ccctag	taaa	ggatgo	cccag	ttgactccgg	gatctcgctt	ccag		44
	<210>	00						
	<211>							
	<211>							
	<213>		sapi	ens				
-+	<400>	99 atasa	~=~=	tgtggtgggg	gaactctggg	tctaatgggg	gagttcatca	60
gcgagc	agge	ctgac	gacac	cacaacaaaa	gaaccccggg	0000009999	9-9	
	<210>	100						
	<211>	57						
	<212>	DNA						
	<213>	Homo	sapi	ens				
	<400>	100						
taacto	tacc	tacco	cagtg	ggatcaccat	gggtccctgt	ctcctccctc	cctccag	57
-99			5 -					
	<210>							
	<211>							
	<212>							
	<213>	Homo	sapı	ens				
	<400>	101						
tctctt	ctcc	tagag	ggcca	tag				23
	<210>	102						
	<211>							
	<212>							
		Homo	sapi	ens				
	<400>	102						
ctatta	cctcc		atete	atgg				24
etgett		ccacc	geece	acgg				
	<210>							
	<211>							
	<212>							
	<213>	Homo	sapı	ens				
	<400>	103						
gctcc	teceg		ctctg	r				20
	J2305	104						
	<210><211>							
	<211>							
		NNA PINA	. eani	ens				

<pre>&lt;400&gt; 104 gggtcccagc caggagcact g 21</pre>
<210> 105 <211> 24 <212> DNA
<213> Homo sapiens
<400> 105 cccctctcac catctcctga tgtg 24
<210> 106 <211> 25
<212> DNA
<213> Homo sapiens
<pre>&lt;400&gt; 106 tggcttcacc ttccctctac ctcgg</pre> 25
<210> 107 <211> 24
<212> DNA
<213> Homo sapiens
<400> 107
tcctttggta ggaaatctag gtgg
<210> 108
<211> 21
<212> DNA <213> Homo sapiens
<pre>&lt;400&gt; 108 ggaagctgga caggcaagag g 21</pre>
<210> 109 <211> 27
<212> DNA
<213> Homo sapiens
<400> 109
atatactgtg ttggaaatct taatgag 27
<210> 110
<211> 21
<212> DNA <213> Homo sapiens
<pre>&lt;400&gt; 110 gctggcacca cagggaatcg g</pre> 21
<210> 111 <211> 25
<211> 23 <212> DNA
<213> Homo sapiens
<400> 111 ctttgcttcc ttgcatcctt ctctg 25
<210> 112
<211> 21 <212> DNA
<212> DNA <213> Homo sapiens
<pre>&lt;400&gt; 112 agccccatg tgcagaatgg g 21</pre>

```
<210> 113
        <211> 21
        <212> DNA
        <213> Homo sapiens
        <400> 113
                                                                                  21
 qqcaqtqatc gagaaacccg g
        <210> 114
        <211> 21
        <212> DNA
        <213> Homo sapiens
        <400> 114
                                                                                  21
 catqccctcc actggggctg g
        <210> 115
        <211> 21
        <212> DNA
        <213> Homo sapiens
        <400> 115
                                                                                  21
 ggatgcccag ttgactccgg g
        <210> 116
        <211> 21
        <212> DNA
        <213> Homo sapiens
        <400> 116
                                                                                  21
 ccccaccaca gtgtcgtcag g
        <210> 117
        <211> 6240
        <212> DNA
        <213> Homo sapiens
        <400> 117
atgctgaggg tcttcatcct ctatgccgag aacgtccaca cacccgacac cgacatcagc
                                                                                 60
                                                                                120
gatgcctact gctccgcggt gtttgcaggg gtgaagaaga gaaccaaagt catcaagaac
agcgtgaacc ctgtatggaa tgagggattt gaatgggacc tcaagggcat ccccctggac
                                                                                180
                                                                                240
cagggetetg agetteatgt ggtggteaaa gaccatgaga cgatggggag gaacaggtte
ctgggggaag ccaaggtccc actccgagag gtcctcgcca cccctagtct gtccgccagc
                                                                                300
ttcaatgccc ccctgctgga caccaagaag cagcccacag gggcctcgct ggtcctgcag gtgtcctaca caccgctgcc tggagctgtg cccctgttcc cgccccctac tcctctggag
                                                                                360
                                                                                420
ccctccccga ctctgcctga cctggatgta gtggcagaca caggaggaga ggaagacaca
                                                                                480
                                                                                540
gaggaccagg gactcactgg agatgaggcg gagccattcc tggatcaaag cggaggcceg
ggggctccca ccaccccaag gaaactacct tcacgtcctc cgccccacta ccccgggatc
                                                                                600
aaaagaaagc gaagtgegec tacatetaga aagetgetgt cagacaaacc geaggattte cagateaggg tecaggtgat egagggege cagetgeegg gggtgaacat caageetgtg
                                                                                660
                                                                                720
gtcaaggtta ccgctgcagg gcagaccaag cggacgcgga tccacaaggg aaacagccca
                                                                                780
                                                                                840
ctcttcaatq agactctttt cttcaacttg tttgactctc ctggggagct gtttgatgag
cccatcttta tcacggtggt agactctcgt tctctcagga cagatgctct cctcggggag
                                                                                900
ttccggatgg acgtgggcac catttacaga gagccccggc acgcctatct caggaagtgg ctgctgctct cagaccctga tgacttctct gctggggcca gaggctacct gaaaacaagc
                                                                                960
                                                                               1020
ctttgtgtgc tggggcctgg ggacgaagcg cctctggaga gaaaagaccc ctctgaagac
                                                                               1080
aaqqaqqaca ttgaaaqcaa cctgctccgg cccacaggcg tagccctgcg aggagcccac
                                                                               1140
ttctgcctga aggtcttccg ggccgaggac ttgccgcaga tggacgatgc cgtgatggac
                                                                               1200
aacgtgaaac agatctttgg cttcgagagt aacaagaaga acttggtgga cccctttgtg
                                                                               1260
gaggtcagct ttgcggggaa aatgctgtgc agcaagatct tggagaagac ggccaacct
                                                                               1320
cagtggaacc agaacatcac actgcctgcc atgtttccct ccatgtgcga aaaaatgagg
                                                                               1380
                                                                               1440
atteqtatea tagactggga cegeetgact cacaatgaca tegtggetac cacetacetg
agtatgtega aaatetetge ceetggagga gaaatagaag aggageetge aggtgetgte aageettega aageeteaga ettggatgae tacetggget teeteeceae ttttgggeee
                                                                               1500
                                                                               1560
tgctacatca acctctatgg cagtcccaga gagttcacag gcttcccaga cccctacaca
                                                                               1620
gageteaaca caggeaaggg ggaaggtgtg gettategtg geeggettet geteteeetg
                                                                               1680
gagaccaagc tggtggagca cagtgaacag aaggtggagg accttcctgc ggatgacatc
                                                                               1740
```

1800

ctccgggtgg agaagtacct taggaggcgc aagtactccc tgtttgcggc cttctactca gccaccatgc tgcaggatgt ggatgatgcc atccagtttg aggtcagcat cgggaactac 1860 gggaacaagt tégacatgae étgécétgéeg etggeéteea céacteagta cagéegtgea 1920 1980 gtetttgacg ggtgccacta ctactaccta ccctggggta acgtgaaacc tgtggtggtg ctgtcatcct actgggagga catcagccat agaatcgaga ctcagaacca gctgcttggg 2040 attgctgacc ggctggaagc tggcctggag caggtccacc tggccctgaa ggcgcagtgc 2100 tccacggagg acgtggactc gctggtggct cagctgacgg atgagctcat cgcaggctgc agccagcctc tgggtgacat ccatgagaca ccctctgcca cccacctgga ccagtacctg 2160 2220 taccagetge geacecatea cetgagecaa ateaetgagg etgeeetgge cetgaagete ggecacagtg ageteetge agetetggag eaggeggagg aetggeteet gegtetgegt geeetggeag aggagececa gaacageetg eeggacateg teatetggat getgeaggga 2280 2340 2400 gacaagegtg tggcatacca gegggtgeee geeeaccaag teetettete eeggeggggt 2460 gccaactact gtggcaagaa ttgtgggaag ctacagacaa tctttctgaa atatccgatg 2520 gagaaggtgc ctggcgcccg gatgccagtg cagatacggg tcaagctgtg gtttgggctc tctgtggatg agaaggagtt caaccagttt gctgagggga agctgtctgt ctttgctgaa 2580 2640 acctatgaga acgagactaa gttggccctt gttgggaact ggggcacaac gggcctcacc 2700 taccccaagt tttctgacgt cacgggcaag atcaagctac ccaaggacag cttccgcccc 2760 2820 teggeegget ggacetggge tggagattgg ttegtgtgte eggagaagae tetgeteeat gacatggacg ccggtcacct gagcttcgtg gaagaggtgt ttgagaacca gacccggctt cccggaggcc agtggatcta catgagtgac aactacaccg atgtgaacgg ggagaaggtg 2880 2940 cttcccaagg atgacattga gtgcccactg ggctggaagt gggaagatga ggaatggtcc 3000 acagacetea acegggetet egatgageaa ggetgggagt atageateae cateceeeg 3060 3120 gagoggaage cgaagcactg ggtccctgct gagaagatgt actacacaca ccgacggcgg egetgggtge geetgegeag gagggatete agecaaatgg aageactgaa aaggeacagg caggeggagg eggagggega gggetgggag taegeetete tttttggetg gaagtteeae etegagtace geaagacaga tgeetteege egeegeegt ggegeegteg catggageea 3180 3240 3300 ctggagaaga cggggcctgc agctgtgttt gcccttgagg gggccctggg cggcgtgatg 3360 3420 gatgacaaga gtgaagatte catgteegte tecacettga getteggtgt gaacagacee acgattteet geatattega etatgggaae egetaceate taegetgeta catgtaceag geeegggaee tggetgegat ggacaaggae tettttetg atecetatge categtetee 3480 3540 ttcctgcacc agagccagaa gacggtggtg gtgaagaaca cccttaaccc cacctgggac 3600 cagacgetea tettetacga gategagate tttggegage eggecacagt tgetgageaa 3660 ccgcccagca ttgtggtgga gctgtacgac catgacactt atggtgcaga cgagtttatg 3720 ggtcgctgca tctgtcaacc gagtctggaa cggatgccac ggctggcctg gttcccactg acgagggca gccagccgtc gggggagctg ctggcctctt ttgagctcat ccagagagag aagccggcca tccaccatat tcctggtttt gaggtgcagg agacatcaag gatcctggat 3780 3840 3900 gagtetgagg acacagacet geectaceca ecaceccaga gggaggeeaa catetacatg 3960 4020 gttcctcaga acatcaagcc agcgctccag cgtaccgcca tcgagatcct ggcatggggc 4080 ctgcggaaca tgaagagtta ccagctggcc aacatctcct cccccagcct cgtggtagag tgtgggggcc agacggtgca gtcctgtgtc atcaggaacc tccggaagaa ccccaacttt 4140 gacatetgea cectetteat ggaagtgatg etgeccaggg aggageteta etgeccece 4200 atcaccgica aggicatcga taaccgccag titiggccgcc ggccigtggt gggccagtgt 4260 accatecget ecetggagag ettectgtgt gaccectaet eggeggagag tecatecea 4320 cagggtggcc cagacgatgt gagcctactc agtcctgggg aagacgtgct catcgacatt gatgacaagg agccctcat cccatccag gaggaagagt tcatcgattg gtggagcaaa ttctttgcct ccataggga gagggaaaag tgcggctcct acctggagaa ggattttgac 4380 4440 4500 accetgaagg tetatgacae acagetggag aatgtggagg cetttgaggg cetgtetgae 4560 ttttgtaaca ccttcaagct gtaccggggc aagacgcagg aggagacaga agatccatct gtgattggtg aatttaaggg cctcttcaaa atttatcccc tcccagaaga cccagccatc 4620 4680 4740 cccatgccc caagacagtt ccaccagctg gccgcccagg gaccccagga gtgcttggtc cgtatctaca ttgtccgagc atttggcctg cagcccaagg accccaatgg aaagtgtgat 4800 ccttacatca agatctccat agggaagaaa tcagtgagtg accaggataa ctacatcccc 4860 4920 tgcacgctgg agcccgtatt tggaaagatg ttcgagctga cctgcactct gcctctggag 4980 aaggacctaa agatcactct ctatgactat gacctcctct ccaaggacga aaagatcggt gagacggtcg tcgacctgga gaacaggctg ctgtccaagt ttggggctcg ctgtggactc ccacagacct actgtgtctc tggaccgaac cagtggcggg accagetccg cccctcccag 5040 5100 5160 ctcctccacc tcttctgcca gcagcataga gtcaaggcac ctgtgtaccg gacagaccgt gtaatgtttc aggataaaga atattccatt gaagagatag aggctggcag gatcccaaac ccacactgg gcccagtgga ggagcgtctg gctctgcatg tgcttcagca gcagggcctg gtcccggagc acgtggagtc acggccctc tacagccccc tgcagccaga catcgagcag 5220 5280 5340 gggaagetge agatgtgggt cgacetattt ccgaaggeec tggggeggec tggaceteec 5400 ttcaacatca ccccacggag agccagaagg tttttcctgc gttgtattat ctggaatacc 5460 5520 agagatgtga tcctggatga cctgagcctc acgggggaga agatgagcga catttatgtg aaaggttgga tgattggctt tgaagaacac aagcaaaaga cagacgtgca ttatcgttcc 5580 ctgggaggtg aaggcaactt caactggagg ttcattttcc ccttcgacta cctgccagct 5640 gagcaagtet gtaccattge caagaaggat geettetgga ggetggacaa gaetgagage 5700 aaaatcccag cacgagtggt gttccagatc tgggacaatg acaagttctc ctttgatgat 5760

tttctggct ccctgcagct cgatctcaac cgcatgcca agccagccaa gacagccaag aagtgctcct tggaccagct ggatgatgct ttccaccag aatggtttgt gtcccttttt gagcagaaaa cagtgaaggg ctggtggccc tgtgtagcag aagagggtga gaagaaaata ctggcgggca agctggaaat gaccttggag attgtagcag agagtgagca tgaggagcgg cccgacacct ccttcctgtg gtttacctcc ccatacaaga ccatgaagtt catcctgtgg gccatcttca tctacgcctt cccgaactat gctgccatga agctggtgaa gcccttcagc	5820 5880 5940 6000 6060 6120 6180 6240
<210> 118	
<211> 13 <212> DNA	
<213> Homo sapiens	
<400> 118 cgcaagcatg ctg	13
<210> 119	
<211> 12 <212> DNA	
<213> Homo sapiens	
<400> 119	12
gagacgatgg gg	
<210> 120 <211> 21	
<211> 21 <212> DNA	
<213> Homo sapiens	
<400> 120	21
gatctaaccc tgctgctcac c	21
<210> 121	
<211> 21 <212> DNA	
<213> Homo sapiens	
<400> 121	
ctggtgtgtt gcagagcgct g	21
<210> 122	
<211> 21	
<212> DNA <213> Homo sapiens	
<400> 122	21
cetetettet getgtettea g	-
<210> 123	
<211> 21 <212> DNA	
<213> Homo sapiens	
<400> 123	
tgtgtctggt tcaccttcgt g	. 21
<210> 124	
<211> 21	
<212> DNA	
<213> Homo sapiens	
<400> 124	21
tccaaataga aatgcctgaa c	
<210> 125 <211> 21	

	.2> DNA .3> Homo sa	apiens			
-10	0> 125				
	c tccaagt	gtt g			21
	.0> 126				
	.1> 21				
	.2> DNA	•			
	.3> Homo sa	apiens			
	00> 126 c agagete	cct g			21
	10> 127				
	1> 19 .2> DNA				
	3> Homo s	apiens			
<40	00> 127				
	g tgctctt	gg			19
_					
	10> 128				
	11> 20				
	L2> DNA L3> Homo s	aniens			
<b>\2</b> .	L3/ HOMO B	aprens			
	00> 128				20
ggagaatt	gc ttgaacc	cag			20
<2	10> 129				
	11> 22				
<2	12> DNA				
<2	13> Homo s	apiens			
<40	00> 129				
	ga tgttgaa	cat tt			22
-2	10> 130				
_	11> 21				
	12> DNA				
	13> Homo s	apiens			
-4	00> 130				
	00> 130 ag cggcgcc	etca a			21
guoodaaa	-9 -99-9				
	10> 131				
	11> 21			•	
	12> DNA				
<2	13> Homo s	sapiens			
<4	00> 131				1
gaccccgg	cg agggtgg	gtcg g			21
-2	10> 132				
<2	11> 24				
	12> DNA				
	13> Homo s	sapiens			
< A	00> 132				
	ca ttctcc	cttt tgtg			24
_		_ ~			
	10> 133				
	11> 24				
	12> DNA 13> Homo a	saniens			
~2	HOMO I				

<400> 133 aggacactgc tgagaaggca cctc	24
<210> 134 <211> 21	
<212> DNA <213> Homo sapiens	
<400> 134 agtgccctgg tggcacgaag g	21
<210> 135 <211> 24	
<212> DNA <213> Homo sapiens	
<400> 135	24
cctacctgca ccttcaagcc atgg	24
<210> 136 <211> 23	
<212> DNA <213> Homo sapiens	
<400> 136 cagaagagcc agggtgcctt agg	23
<210> 137	
<211> 24	
<212> DNA <213> Homo sapiens	
<400> 137	
ccttggacct taacctggca gagg	24
<210> 138 <211> 21	
<211> 21 <212> DNA	
<213> Homo sapiens	
<400> 138	21
cgaggccagc gcaccaacct g	21
<210> 139	
<211> 22 <212> DNA	
<213> Homo sapiens	
<400> 139	22
actgccggcc attcttgctg gg	22
<210> 140	
<211> 21	
<212> DNA <213> Homo sapiens	
<400> 140	
ccaggcctca ttagggccct c	21
<210> 141 <211> 22	
<212> DNA	
<213> Homo sapiens	
<400> 141	22
ctqaaqagqa gcctggggtc ag	22

	<210> 142	
	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
	<400> 142	24
ctgaga	atttc tgactcttgg ggtg	24
	<210> 143	
	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
	4005 142	
	<pre>&lt;400&gt; 143 tctgc cctcatgccc catg</pre>	24
aaggu	construction carry	
	<210> 144	
	<211> 21	
	<212> DNA	
	<213> Homo sapiens	
	<400> 144	21
ctggc	ctgag ggatcagcag g	21
	2.2. 4.2	
	<210> 145	
	<211> 23	
	<212> DNA	
	<213> Homo sapiens	
	<400> 145	
ataca	tacat acageceaeg gag	23
90900		
	<210> 146	
	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
	<400> 146	24
gagct	attgg gttggccgtg tggg	
	<210> 147	
	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
	<400> 147	
accaa	cacgg agaagtgaga actg	24
	<210> 148	
	<211> 26	
	<212> DNA	
	<213> Homo sapiens	
	<400> 148	
ccaca	acttta tttaacgctt tggcgg	26
CCACA		
	<210> 149	
	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
	<400> 149	24
cagaa	accaaa atgcaaggat acgg	
	-210> 150	
	<210> 150 <211> 25	
	<211> 25 <212> DNA	
	78 687 8777	

<21	3> Homo	sanie	ns			
72.2	1201110	Jupio				
<40	0> 150					
cttctgatt	c tgggat	cacc	aaagg			25
-01	05 151					
	0> 151					
	1> 22					
	2> DNA		~~			
<21	13> Homo	sapie	ns			
<40	00> 151					
ggaccgtaa		eccaq	aa			22
gguccgca	9 9		33			
	152					
	11> 24					
<21	l2> DNA					
<21	L3> Homo	sapie	ns			
- 40	305 152					
	00> 152					24
cctgtgct	a ggagc	gcatg	aagg			
<21	10> 153					
	11> 22					
	L2> DNA					
	L3> Homo	sapie	ns			
<40	00> 153					
gcagacct	cc cacco	aaggg	cg			22
	10> 154					
	11> 24					
	12> DNA					
<2.	13> Homo	Bapie	ens			
<40	00> 154					
gagacagat	tg gggga	cagtc	aggg			24
	10> 155					
	11> 21					
	12> DNA					
<2	13> Homo	sapie	ens			
- 41	00> 155					
cctcccga		ctect	α			21
coccega	ga gaacc		9			
<2.	10> 156					
<2	11> 21					
<2	12> DNA					
<2	13> Homo	sapie	ens			
الله م	005 156					
	00> 156		~	•		21
gggagccc	ag agtee	ccatg	g			
<2	10> 157					
	11> 21					
	12> DNA					
	13> Homo	sapie	ens			
		•				
	00> 157			•		٠.
gggcctcc	tt gggtt	tgctg	g			2:
-2	10> 158					
	11> 21				•	
	11> 21 12> DNA					
	13> Homo	ganie	ens			
~~						

	<400> 158	21
gcctcc	ccag catcctgccg g	
	<210> 159	
	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
	<400> 159	24
tcactg	agcc gaatgaaact gagg	
	<210> 160	
	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
	<400> 160	
	ctga gttcctttcc tgtg	24
cgcggc	cega gaccodado egraj	
	<210> 161	
	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
	-400> 161	
	<400> 161 aggg cagaacgaag aggg	24
ggccaa	aggg cagaacgaag aggg	
	<210> 162	
	<211> 21	
	<212> DNA	
	<213> Homo sapiens	
	.400. 160	
	<400> 162 ecttc teccagecat g	21
ceegro	ecte teceagecat g	
	<210> 163	
	<211> 21	
	<212> DNA	
	<213> Homo sapiens	
	<400> 163	21
CTCCCC	etggt tgtccccaag g	
	<210> 164	
	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
	<400> 164	24
cgacco	ectet gattgecact tgtg	
	<210> 165	
	<211> 21	
	<212> DNA	
	<213> Homo sapiens	
	<400> 165	2:
ggcato	cctgc ccttgccagg g	-
	<210> 166	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
<b>.</b>	<pre>&lt;400&gt; 166 ctccc ctqctccttq</pre>	20
CCECT	CLUCC COUCLOS	

	<210> 167	
	<211> 21	
	<212> DNA	
	<213> Homo sapiens	
	_	
	<400> 167	21
cttcc	ctgcc ccgacgccca g	21
	<210> 168	
	<211> 21	
	<212> DNA	
	<213> Homo sapiens	
	CZ13> HOMO Baptens	
	100, 100	
	<400> 168	21
cagcg	ctcag gcccgtctct c	
	<210> 169	
	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
	<400> 169	
		24
tgcat	aggca tgtgcagctt tggg	
	<210> 170	
	<211> 21	
	<212> DNA	
	<213> Homo sapiens	
	-	
	<400> 170	
catoo	accct ctgccctgtg g	21
cacgo	acced edge-edge-gray	
	<210> 171	
	<211> 21	
	<212> DNA	
	<213> Homo sapiens	
	<400> 171	21
agttg	rageca ggagaggtgg g	2.1
	<210> 172	
	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
	<400> 172	
		24
catca	aggogo attocatotg toog	
	1010 173	
	<210> 173	
	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
	<400> 173	~ -
agcag	ggagag cagaagaaga aagg	24
•		
	<210> 174	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
	12107 HOMO Baptone	
	ZADON 174	
	<400> 174	22
gtgt	gtcacc atccccaccc cg	
	<210> 175	
	<211> 25	
	<212> DNA	

<213> Homo sapiens	
<400> 175	25
caagagatgg gagaaaggcc ttatg	25
<210> 176	
<211> 23	
<212> DNA	
<213> Homo sapiens	
<400> 176	23
ctgggacatc cggatcctga agg	20
<210> 177	
<211> 22	
<212> DNA	
<213> Homo sapiens	
<400> 177	22
tccaggtagt gggaggcaga gg	22
<210> 178	
<211> 24	
<212> DNA	
<213> Homo sapiens	
<400> 178	
tcccactacc tggagctgcc ttgg	24
<210> 179	
<211> 21	
<212> DNA	
<213> Homo sapiens	
<400> 179	
ggctctcccc agccctccct g	21
<210> 180	
<211> 24	
<212> DNA	
<213> Homo sapiens	
<400> 180	24
cagagcagca gagactctga ccag	24
<210> 181	
<211> 21	
<212> DNA	
<213> Homo sapiens	
<400> 181	21
tagaccccac ctgcccctga g	21
<210> 182	
<211> 24	
<212> DNA	
<213> Homo sapiens	
<400> 182	24
teeteteatt gettgeetgt tegg	24
<210> 183	
<211> 21 <212> DNA	
<212> DNA <213> Homo ganiens	

<400> 183 ttgagagctt gccggggatg g	21
<210> 184 <211> 24 <212> DNA	
<213> Homo sapiens	
<400> 184 aagtgccaag caatgagtga ccgg	24
<210> 185 <211> 21	
<212> DNA <213> Homo sapiens	
<400> 185	
ctcactccca cccaccacct g	21
<210> 186 <211> 21 <212> DNA	
<212> DNA <213> Homo sapiens	
<400> 186	21
cccaccggcc tctgagtctg c	
<210> 187 <211> 24	
<212> DNA	
<213> Homo sapiens	
<400> 187 accctaccca agccaggaca agtg	24
<210> 188	
<211> 24 <212> DNA	
<213> Homo sapiens	
<400> 188	24
gaatctgcca taaccagctt cgtg	2.4
<210> 189	
<211> 24 <212> DNA	
<213> Homo sapiens	
<400> 189	2.4
tatcacccca tagaggcctc gaag	24
<210> 190	
<211> 24 <212> DNA	
<213> Homo sapiens	
<pre>&lt;400&gt; 190 cagccactca ctctggcacc tctg</pre>	24
<210> 191	
<211> 24	
<212> DNA <213> Homo sapiens	
<400> 191	24
agcccacagt ctctgactct cctg	24

<210> 192 <211> 24 <212> DNA <213> Homo sapiens	
<400> 192 acatetetea gggteeetge tgtg	24
<210> 193	
<211> 21	
<212> DNA <213> Homo sapiens	
<400> 193	
cctgtgaggg gacgaggcag g	21
<210> 194	
<211> 24 <212> DNA	
<213> Homo sapiens	
<400> 194	24
gccctgggta agggatgctg attc	
<210> 195	
<211> 21 <212> DNA	
<213> Homo sapiens	
<400> 195	21
cctgcctggg cctcctggat c	
<210> 196	
<211> 21 <212> DNA	
<213> Homo sapiens	
<400> 196	21
gagggtgatg ggggccttag g	
<210> 197	
<211> 24 <212> DNA	
<213> Homo sapiens	
<400> 197	24
gcaatcagtt tgaagaagga aagg	
<210> 198	
<211> 24 <212> DNA	
<213> Homo sapiens	
<400> 198	24
cacctttgtc tccattctac ctgc	
<210> 199	
<211> 21 <212> DNA	
<213> Homo sapiens	
<400> 199	21
ctcccagccc ccacgcccag g	21
<210> 200	
<211> 24 <212> DNA	

<213> Homo sapiens	
<400> 200	
ctgagccact ctcctcattc tgtg	24
<210> 201	
<211> 21	
<212> DNA	
<213> Homo sapiens	
<400> 201	
tggaagggga cagtagggag g	21
<210> 202	
<211> 22	
<212> DNA	
<213> Homo sapiens	
<400> 202	
ggccagtgcg ttcttcctcc tc	22
<210> 203	
<211> 22	
<212> DNA	
<213> Homo sapiens	
<400> 203	
tecetgacet geceateate te	22
<210> 204	
<211> 21	
<212> DNA	
<213> Homo sapiens	
<400> 204	
gccctgtca ggcctggatg g	21
<210> 205	
<210> 205 <211> 21	
<211> 21 <212> DNA	
<213> Homo sapiens	
<400> 205	21
tgacccaggc ctccctggag g	
<210> 206	
<211> 24	
<212> DNA	
<213> Homo sapiens	
<400> 206	
ctgaaatggt ctctttcttt ctac	24
<210> 207	
<211> 24	
<212> DNA	
<213> Homo sapiens	
<400> 207	
cacaccgact gtcagactga agag	24
<210> 208	
<211> 24	
<212> DNA	
<213> Homo sapiens	

<400> 208	0.4
ttgtccctc ctctaatccc catg	24
Ligitation deducation and	
2210× 200	
<210> 209	
<211> 21	
<212> DNA	
<213> Homo sapiens	
<400> 209	0.1
gggttaggga cgtcttcgag g	21
gggccaggga egerress.s	
<210> 210	
<210> 210 <211> 22	
<212> DNA	
<213> Homo sapiens	
<400> 210	22
cagccaaacc atatcaacaa tg	22
•	
<210> 211	
<211> 21	
<211> 21 <212> DNA	
(212) DNA	
<213> Homo sapiens	
<400> 211	21
ctggggaggt gagggctcta g	2.1
<210> 212	
<211> 21	
<212> DNA	
<213> Homo sapiens	
<400> 212	21
gaagtgtttt gtctcctcct c	
<210> 213	
<211> 20	
<211> 20 <212> DNA	
<213> Homo sapiens	
<213> Homo Bapiens	
<400> 213	20
gcaggcagcc agcccccatc	
<210> 214	
<211> 21	
<212> DNA	
<213> Homo sapiens	
(213) Homo Baptone	
4400> 014	
<400> 214	21
gggtgccctg tgttggctga c	
_	
<210> 215	
<211> 20	
<212> DNA	
<213> Homo sapiens	
•	
<400> 215	
74007 213	20
gcaggcagcc agcccccatc	
4010> 216	·
<210> 216	
<211> 24	
<212> DNA	
<213> Homo sapiens	
<400> 216	
ctcgtctatg tcttgtgctt gctc	24

<210> 217 <211> 23 <212> DNA <213> Homo sapiens	
<400> 217 caccatggtt tggggtcatg tgg	23
<210> 218 <211> 21 <212> DNA <213> Homo sapiens	
<400> 218 tctcgcttcc ccagctcctg c	21
<210> 219 <211> 22 <212> DNA <213> Homo sapiens	
<400> 219 tctggagttc gaggactctg gg	22
<210> 220 <211> 21 <212> DNA <213> Homo sapiens	
<400> 220 agaagggtgg ggagagaacg g	21
<210> 221 <211> 21 <212> DNA <213> Homo sapiens	
<400> 221 cagctcagag cctgtggctg g	21
<210> 222 <211> 24 <212> DNA <213> Homo sapiens	
<400> 222 aaggcettee cateetttgg tagg	24
<210> 223 <211> 21 <212> DNA <213> Homo sapiens	
<400> 223 acaacccaga gggagcacgg g	21
<210> 224 <211> 25 <212> DNA <213> Homo sapiens	
<400> 224 gttgacgatg tatatactgt gttgg	25
<210> 225 <211> 25 <212> DNA	

<213> Homo sapiens				
<400> 225				
gcctctctct aactttgctt ccttg				25
george addition to the				
<210> 226				
<211> 24				
<212> DNA				
<213> Homo sapiens				
<400> 226				
ggctacaggc tggcagtgat cgag				24
<210> 227				
<211> 21				
<212> DNA				
<213> Homo sapiens				
<400> 227				2.1
ttcccccatg ccctccactg g				21
<210> 228				
<211> 24				
<212> DNA				
<213> Homo sapiens				
<400> 228				24
agccttcgtg cccctaacca agtg				24
<210> 229				
<211> 21				
<212> DNA				
<213> Homo sapiens				
<400× 229				
<400> 229				21
ctgtgggcat tggggctcag g				
<210> 230				
<211> 20				
<212> DNA				
<213> Homo sapiens				
<400> 230				
gccccagtgg gatcaccatg				20
g				
<210> 231				
<211> 21				
<212> DNA				
<213> Homo sapiens				
<400> 231				
atgctggagg ggaccccacg g				21
<210> 232				
<211> 3671				
<212> DNA				
<213> Homo sapiens				
<b>-220</b> \				
<220>	•			
<221> CDS				
<222> (418)(3381)				
<400> 232				
<pre>&lt;400&gt; 232 tcctggttca agcgattctc tggcctcagc (</pre>	ctocoaata	actaggatta	caggeatget	60
ccaccaagee egggtaattt tgtattttta	atagagaggg	gattttacca	tattaatcaa	120
gctggtctcg aactcctgac ctcaggtgat	ctocccacct	tggcctccca	acgtgctgag	180
attacaggea tgagtcactg tgcccggcag	agatggtcta	attcatatga	aagaactctg	240
		-	-	

ctaa	cctt	at o	gaat	tttt	t tt	tttt	gaga	ago	aaat	tgc	aaat	ttgt	ga t	agat	atcac ctaaa ag atg Met 1	300 360 420
ctc ( Leu (	G1y ggg	aga Arg	ttg Leu 5	agc Ser	cta Leu	gag Glu	aaa Lys	gga Gly 10	aga Arg	ttt Phe	gtg Val	aac Asn	cca Pro 15	gga Gly	ggc Gly	468
aga ( Arg	ggt Gly	aga Arg 20	gat Asp	cca Pro	gga Gly	gag Glu	ggc Gly 25	ggc Gly	gtg Val	atg Met	gat Asp	gac Asp 30	aag Lys	agt Ser	gaa Glu	516
gat Asp	tcc Ser 35	atg Met	tcc Ser	gtc Val	tcc Ser	acc Thr 40	ttg Leu	agc Ser	ttc Phe	ggt Gly	gtg Val 45	aac Asn	aga Arg	ccc Pro	acg Thr	564
att Ile 50	tcc Ser	tgc Cys	ata Ile	ttc Phe	gac Asp 55	tat Tyr	ggg Gly	aac Asn	cgc Arg	tac Tyr 60	cat His	cta Leu	cgc Arg	tgc Cys	tac Tyr 65	612
atg Met	tac Tyr	cag Gln	gcc Ala	cgg Arg 70	gac Asp	ctg Leu	gct Ala	gcg Ala	atg Met 75	gac Asp	aag Lys	gac Asp	tct Ser	ttt Phe 80	tct Ser	660
gat Asp	ccc Pro	tat Tyr	gcc Ala 85	atc Ile	gtc Val	tcc Ser	ttc Phe	ctg Leu 90	cac His	cag Gln	agc Ser	cag Gln	aag Lys 95	acg Thr	gtg Val	708
gtg Val	gtg Val	aag Lys 100	aac Asn	acc Thr	ctt Leu	aac Asn	ccc Pro 105	acc Thr	tgg Trp	gac Asp	cag Gln	acg Thr 110	ctc Leu	atc Ile	ttc Phe	756
tac Tyr	gag Glu 115	atc Ile	gag Glu	atc Ile	ttt Phe	ggc Gly 120	gag Glu	ccg Pro	gcc Ala	aca Thr	gtt Val 125	gct Ala	gag Glu	caa Gln	ccg Pro	804
ccc Pro 130	agc Ser	att Ile	gtg Val	gtg Val	gag Glu 135	ctg Leu	tac Tyr	gac Asp	cat His	gac Asp 140	act Thr	tat Tyr	ggt Gly	gca Ala	gac Asp 145	852
gag Glu	ttt Phe	atg Met	ggt Gly	cgc Arg 150	tgc Cys	atc Ile	tgt Cys	caa Gln	ccg Pro 155	agt Ser	ctg Leu	gaa Glu	cgg Arg	atg Met 160	cca Pro	900
cgg Arg	ctg Leu	gcc Ala	tgg Trp 165	ttc Phe	cca Pro	ctg Leu	acg Thr	agg Arg 170	ggc Gly	agc Ser	cag Gln	ccg Pro	tcg Ser 175	ggg Gly	gag Glu	948
ctg Leu	ctg Leu	gcc Ala 180	tct Ser	ttt Phe	gag Glu	ctc Leu	atc Ile 185	cag Gln	aga Arg	gag Glu	aag Lys	ccg Pro 190	gcc Ala	atc Ile	cac His	996
cat His	att Ile 195	cct Pro	ggt Gly	ttt Phe	gag Glu	gtg Val 200	cag Gln	gag Glu	aca Thr	tca Ser	agg Arg 205	atc Ile	ctg Leu	gat Asp	gag Glu	1044
tct Ser 210	gag Glu	gac Asp	aca Thr	gac Asp	ctg Leu 215	ccc Pro	tac Tyr	cca Pro	cca Pro	ccc Pro 220	cag Gln	agg Arg	gag Glu	gcc Ala	aac Asn 225	1092
atc Ile	tac Tyr	atg Met	gtt Val	cct Pro 230	Gln	aac Asn	atc Ile	aag Lys	cca Pro 235	Ala	ctc Leu	cag Gln	cgt Arg	acc Thr 240	Ala	1140

atc Ile	gag Glu	atc Ile	ctg Leu 245	gca Ala	tgg Trp	ggc Gly	ctg Leu	cgg Arg 250	aac Asn	atg Met	aag Lys	agt Ser	tac Tyr 255	cag Gln	ctg Leu	1188
gcc Ala	aac Asn	atc Ile 260	tcc Ser	tcc Ser	ccc Pro	agc Ser	ctc Leu 265	gtg Val	gta Val	gag Glu	tgt Cys	999 Gly 270	ggc Gly	cag Gln	acg Thr	1236
gtg Val	cag Gln 275	tcc Ser	tgt Cys	gtc Val	atc Ile	agg Arg 280	aac Asn	ctc Leu	cgg Arg	aag Lys	aac Asn 285	ccc Pro	aac Asn	ttt Phe	gac Asp	1284
atc Ile 290	tgc Cys	acc Thr	ctc Leu	ttc Phe	atg Met 295	gaa Glu	gtg Val	atg Met	ctg Leu	ccc Pro 300	agg Arg	gag Glu	gag Glu	ctc Leu	tac Tyr 305	1332
tgc Cys	ccc Pro	ccc Pro	atc Ile	acc Thr 310	gtc Val	aag Lys	gtc Val	atc Ile	gat Asp 315	aac Asn	cgc Arg	cag Gln	ttt Phe	ggc Gly 320	cgc Arg	1380
cgg Arg	cct Pro	gtg Val	gtg Val 325	ggc Gly	cag Gln	tgt Cys	acc Thr	atc Ile 330	cgc Arg	tcc Ser	ctg Leu	gag Glu	agc Ser 335	ttc Phe	ctg Leu	1428
tgt Cys	gac Asp	ccc Pro 340	tac Tyr	tcg Ser	gcg Ala	gag Glu	agt Ser 345	cca Pro	tcc Ser	cca Pro	cag Gln	ggt Gly 350	ggc Gly	cca Pro	gac Asp	1476
gat Asp	gtg Val 355	agc Ser	cta Leu	ctc Leu	agt Ser	cct Pro 360	ggg Gly	gaa Glu	gac Asp	gtg Val	ctc Leu 365	atc Ile	yab dac	att Ile	gat Asp	1524
gac Asp 370	aag Lys	gag Glu	ccc Pro	ctc Leu	atc Ile 375	ccc Pro	atc Ile	cag Gln	gag Glu	gaa Glu 380	gag Glu	ttc Phe	atc Ile	gat Asp	tgg Trp 385	1572
tgg Trp	agc Ser	aaa Lys	ttc Phe	ttt Phe 390	gcc Ala	tcc Ser	ata Ile	ggg Gly	gag Glu 395	agg Arg	gaa Glu	aag Lys	tgc Cys	ggc Gly 400	tcc Ser	1620
tac Tyr	ctg Leu	gag Glu	aag Lys 405	gat Asp	ttt Phe	gac Asp	acc Thr	ctg Leu 410	aag Lys	gtc Val	tat Tyr	gac Asp	aca Thr 415	cag Gln	ctg Leu	1668
gag Glu	aat Asn	gtg Val 420	Glu	gcc Ala	ttt Phe	gag Glu	ggc Gly 425	Leu	tct Ser	gac Asp	ttt Phe	tgt Cys 430	aac Asn	acc Thr	ttc Phe	1716
aag Lys	ctg Leu 435	Tyr	cgg Arg	ggc Gly	aag Lys	acg Thr 440	Gln	gag Glu	gag Glu	aca Thr	gaa Glu 445	gat Asp	cca Pro	tct Ser	gtg Val	1764
att Ile 450	Gly	gaa Glu	ttt Phe	aag Lys	ggc Gly 455	Leu	ttc Phe	aaa Lys	att Ile	tat Tyr 460	ccc Pro	ctc Leu	cca Pro	gaa Glu	gac Asp 465	1812
cca Pro	gcc Ala	ato	ccc Pro	atg Met 470	Pro	cca Pro	aga Arg	cag Gln	ttc Phe 475	His	cag Gln	ctg Leu	gcc	gcc Ala 480	GIn	1860
gga Gly	ccc Pro	cag Gln	gag Glu 485	Сув	ttg Leu	gtc Val	cgt Arg	atc Ile 490	Tyr	att Ile	gtc Val	cga Arg	gca Ala 495	Pne	ggc	1908
ctg Leu	g cag Gln	ccc Pro	) Lye	gac Asp	ccc Pro	aat Asn	gga Gly 505	Lys	tgt Cys	gat Asp	cct Pro	tac Tyr 510	. TTE	aag Lys	atc Ile	1956

tcc Ser	ata Ile 515	ggg Gly	aag Lys	aaa Lys	tca Ser	gtg Val 520	agt Ser	gac Asp	cag Gln	gat Asp	aac Asn 525	tac Tyr	atc Ile	ccc Pro	tgc Cys	2004
acg Thr 530	ctg Leu	gag Glu	ccc Pro	gta Val	ttt Phe 535	gga Gly	aag Lys	atg Met	ttc Phe	gag Glu 540	ctg Leu	acc Thr	tgc Cys	act Thr	ctg Leu 545	2052
cct Pro	ctg Leu	gag Glu	aag Lys	gac Asp 550	cta Leu	aag Lys	atc Ile	act Thr	ctc Leu 555	tat Tyr	gac Asp	tat Tyr	gac Asp	ctc Leu 560	ctc Leu	2100
tcc Ser	aag Lys	gac Asp	gaa Glu 565	aag Lys	atc Ile	ggt Gly	gag Glu	acg Thr 570	gtc Val	gtc Val	gac Asp	ctg Leu	gag Glu 575	aac Asn	agg Arg	2148
ctg Leu	ctg Leu	tcc Ser 580	aag Lys	ttt Phe	ggg ggg	gct Ala	cgc Arg 585	tgt Cys	gga Gly	ctc Leu	cca Pro	cag Gln 590	acc Thr	tac Tyr	tgt Cys	2196
gtc Val	tct Ser 595	gga Gly	ccg Pro	aac Asn	cag Gln	tgg Trp 600	cgg Arg	gac Asp	cag Gln	ctc Leu	cgc Arg 605	ccc Pro	tcc Ser	cag Gln	ctc Leu	2244
ctc Leu 610	cac His	ctc Leu	ttc Phe	tgc Cys	cag Gln 615	cag Gln	cat His	aga Arg	gtc Val	aag Lys 620	gca Ala	cct Pro	gtg Val	tac Tyr	cgg Arg 625	2292
aca Thr	gac Asp	cgt Arg	gta Val	atg Met 630	ttt Phe	cag Gln	gat Asp	aaa Lys	gaa Glu 635	tat Tyr	tcc Ser	att Ile	gaa Glu	gag Glu 640	ata Ile	2340
gag Glu	gct Ala	ggc Gly	agg Arg 645	atc Ile	cca Pro	aac Asn	cca Pro	cac His 650	ctg Leu	ggc Gly	cca Pro	gtg Val	gag Glu 655	gag Glu	cgt Arg	2388
ctg Leu	gct Ala	ctg Leu 660	cat His	gtg Val	ctt Leu	cag Gln	cag Gln 665	cag Gln	ggc Gly	ctg Leu	gtc Val	ccg Pro 670	gag Glu	cac His	gtg Val	2436
gag Glu	tca Ser 675	cgg Arg	ccc Pro	ctc Leu	tac Tyr	agc Ser 680	ccc Pro	ctg Leu	cag Gln	cca Pro	gac Asp 685	atc Ile	gag Glu	cag Gln	ggg ggg	2484
aag Lys 690	Leu	Gln	atg Met	Trp	Val	Asp	Leu	Phe	Pro	Lys	Ala	ctg Leu	ggg Gly	cgg Arg	cct Pro 705	2532
gga Gly	cct Pro	ccc Pro	ttc Phe	aac Asn 710	atc Ile	acc Thr	cca Pro	cgg Arg	aga Arg 715	gcc Ala	aga Arg	agg Arg	ttt Phe	ttc Phe 720	ctg Leu	2580
cgt Arg	tgt Cys	att Ile	atc Ile 725	tgg Trp	aat Asn	acc Thr	aga Arg	gat Asp 730	gtg Val	atc Ile	ctg Leu	gat Asp	gac Asp 735	ctg Leu	agc Ser	2628
ctc Leu	acg Thr	ggg Gly 740	gag Glu	aag Lys	atg Met	agc Ser	gac Asp 745	att Ile	tat Tyr	gtg Val	aaa Lys	ggt Gly 750	Trp	atg Met	att Ile	2676
ggc Gly	ttt Phe 755	Glu	gaa Glu	cac His	aag Lys	caa Gln 760	Lys	aca Thr	gac Asp	gtg Val	cat His 765	tat Tyr	cgt Arg	tcc Ser	ctg Leu	2724

									•							
gga Gly 770	ggt Gly	gaa Glu	ggc Gly	aac Asn	ttc Phe 775	aac Asn	tgg Trp	agg Arg	ttc Phe	att Ile 780	ttc Phe	ccc Pro	ttc Phe	gac Asp	tac Tyr 785	2772
ctg Leu	cca Pro	gct Ala	gag Glu	caa Gln 790	gtc Val	<b>Cys</b>	acc Thr	att Ile	gcc Ala 795	aag Lys	aag Lys	gat Asp	gcc Ala	ttc Phe 800	tgg Trp	2820
agg Arg	ctg Leu	gac Asp	aag Lys 805	act Thr	gag Glu	agc Ser	aaa Lys	atc Ile 810	cca Pro	gca Ala	cga Arg	gtg Val	gtg Val 815	ttc Phe	cag Gln	2868
atc Ile	tgg Trp	gac Asp 820	aat Asn	gac Asp	aag Lys	ttc Phe	tcc Ser 825	ttt Phe	gat Asp	gat Asp	ttt Phe	ctg Leu 830	ggc Gly	tcc Ser	ctg Leu	2916
cag Gln	ctc Leu 835	gat Asp	ctc Leu	aac Asn	cgc Arg	atg Met 840	ccc Pro	aag Lys	cca Pro	gcc Ala	aag Lys 845	aca Thr	gcc Ala	aag Lys	aag Lys	2964
tgc Cys 850	tcc Ser	ttg Leu	gac Asp	cag Gln	ctg Leu 855	gat Asp	gat Asp	gct Ala	ttc Phe	cac His 860	cca Pro	gaa Glu	tgg Trp	ttt Phe	gtg Val 865	3012
tcc Ser	ctt Leu	ttt Phe	gag Glu	cag Gln 870	aaa Lys	aca Thr	gtg Val	aag Lys	ggc Gly 875	tgg Trp	tgg Trp	ccc Pro	tgt Cys	gta Val 880	gca Ala	3060
gaa Glu	gag Glu	ggt Gly	gag Glu 885	aag Lys	aaa Lys	ata Ile	ctg Leu	gcg Ala 890	ggc Gly	aag Lys	ctg Leu	gaa Glu	atg Met 895	acc Thr	ttg Leu	3108
gag Glu	att Ile	gta Val 900	Ala	gag Glu	agt Ser	gag Glu	cat His 905	gag Glu	gag Glu	cgg Arg	cct Pro	gct Ala 910	GIY	cag Gln	ggc Gly	3156
cgg Arg	gat Asp 915	gag Glu	ccc Pro	aac Asn	atg Met	aac Asn 920	Pro	aag Lys	ctt Leu	gag Glu	gac Asp 925	Pro	agg Arg	cgc Arg	ccc Pro	3204
gac Asp 930	Thr	tcc Ser	ttc Phe	ctg Leu	tgg Trp 935	Phe	acc Thr	tcc Ser	cca Pro	tac Tyr 940	гÃа	acc Thr	atg Met	aag Lys	ttc Phe 945	3252
atc Ile	ctg Leu	tgg Trp	cgg Arg	cgt Arg 950	Phe	cgg Arg	tgg Trp	gcc Ala	atc Ile 955	TIE	ctc Leu	tto Phe	ato Ile	atc 1le 960	ctc Leu	3300
ttc Phe	ato Ile	ctç Leu	ctg Leu 965	Leu	tto Phe	ctg Leu	gcc	ato Ile 970	Phe	atc Ile	tac Tyr	gcc Ala	tto Phe 975	Pro	aac Asn	3348
tat Tyr	gct Ala	gco Ala 980	a Met	aaç Lys	ctg Lev	gtg Val	aag Lys 985	Pro	ttc Phe	ago Ser	tga	iggad	tct	cctg	jecetgt	3401
cga gad	cggc	ctc	caga	ccto ccca caaco	ect a aga c eca a	iggco ittgo icgct	tgat taac tttt	t gt a to t go	ccto	ccag	g ggt	gggc	ccc	actt	igctcgg icagatg iccatca iataaaa	3461 3521 3581 3641 3671

<210> 233 <211> 988 <212> PRT <213> Homo sapiens

<400> 233 Met Leu Gly Arg Leu Ser Leu Glu Lys Gly Arg Phe Val Asn Pro Gly Gly Arg Gly Arg Asp Pro Gly Glu Gly Gly Val Met Asp Asp Lys Ser Glu Asp Ser Met Ser Val Ser Thr Leu Ser Phe Gly Val Asn Arg Pro Thr Ile Ser Cys Ile Phe Asp Tyr Gly Asn Arg Tyr His Leu Arg Cys Tyr Met Tyr Gln Ala Arg Asp Leu Ala Ala Met Asp Lys Asp Ser Phe Ser Asp Pro Tyr Ala Ile Val Ser Phe Leu His Gln Ser Gln Lys Thr Val Val Val Lys Asn Thr Leu Asn Pro Thr Trp Asp Gln Thr Leu Ile Phe Tyr Glu Ile Glu Ile Phe Gly Glu Pro Ala Thr Val Ala Glu Gln Pro Pro Ser Ile Val Val Glu Leu Tyr Asp His Asp Thr Tyr Gly Ala Asp Glu Phe Met Gly Arg Cys Ile Cys Gln Pro Ser Leu Glu Arg Met Pro Arg Leu Ala Trp Phe Pro Leu Thr Arg Gly Ser Gln Pro Ser Gly Glu Leu Leu Ala Ser Phe Glu Leu Ile Gln Arg Glu Lys Pro Ala Ile His His Ile Pro Gly Phe Glu Val Gln Glu Thr Ser Arg Ile Leu Asp Glu Ser Glu Asp Thr Asp Leu Pro Tyr Pro Pro Pro Gln Arg Glu Ala Asn Ile Tyr Met Val Pro Gln Asn Ile Lys Pro Ala Leu Gln Arg Thr Ala Ile Glu Ile Leu Ala Trp Gly Leu Arg Asn Met Lys Ser Tyr Gln Leu Ala Asn Ile Ser Ser Pro Ser Leu Val Val Glu Cys Gly Gly Thr Val Gln Ser Cys Val Ile Arg Asn Leu Arg Lys Asn Pro Asn Phe Asp Ile Cys Thr Leu Phe Met Glu Val Met Leu Pro Arg Glu Glu Leu Tyr Cys Pro Pro Ile Thr Val Lys Val Ile Asp Asn Arg Gln Phe Gly Arg Arg Pro Val Val Gly Gln Cys Thr Ile Arg Ser Leu Glu Ser Phe Leu Cys Asp Pro Tyr Ser Ala Glu Ser Pro Ser Pro Gln Gly Gly Pro Asp Asp Val Ser Leu Leu Ser Pro Gly Glu Asp Val Leu Ile Asp Ile Asp Asp Lys Glu Pro Leu Ile Pro Ile Gln Glu Glu Glu Phe Ile Asp Trp Trp Ser Lys Phe Phe Ala Ser Ile Gly Glu Arg Glu Lys Cys Gly Ser Tyr Leu Glu Lys Asp Phe Asp Thr Leu Lys Val Tyr Asp Thr Gln Leu Glu Asn Val Glu Ala Phe Glu Gly Leu Ser Asp Phe Cys Asn Thr Phe Lys Leu Tyr Arg Gly Lys Thr Gln Glu Glu Thr Glu Asp Pro Ser Val Ile Gly Glu Phe Lys Gly Leu Phe Lys Ile Tyr Pro Leu Pro Glu Asp Pro Ala Ile Pro Met Pro Pro Arg Gln Phe His Gln Leu Ala Ala Gln Gly Pro Gln Glu Cys Leu Val Arg Ile Tyr Ile Val Arg Ala Phe Gly Leu Gln Pro Lys Asp Pro Asn Gly Lys Cys Asp Pro Tyr Ile Lys Ile Ser Ile Gly Lys Lys Ser Val Ser Asp Gln Asp Asn Tyr Ile Pro 

												_		<b>~</b>	m b
_			Glu			E 2 L					340				
	Pro		Glu												
Leu			Asp						3,0						
			Ser 580	Lys				כסכ					J J U		
_		EOE	Gly				600					000			
	C 1 A	His	Leu			615					02 U				
C 0 F	Thr		Arg		630					0.00					
Ile			Gly	C 1 E					uca					0,00	
_			Leu 660					כמם					0,0		
		<i>~~</i>	Arg				กกบ					000			
	$\epsilon \alpha \alpha$	Leu	Gln			645					,,,,,				
205	Gly		Pro		710					110					, 2.0
Leu			Ile	725					7.30					, , ,	
			Gly 740					743					, , ,		
		755	Glu				760					,03			
			Glu			775					, 60				
	Leu		Ala		7,011					,,,,					
Trp			Asp	ONE					010						
			Asp 820					nz o	1						
		025	Asp				840	1				043			
	OFA		Leu			* ~ ~					000				
~~-			Phe		0.77					0/2					
Ala	Glu		Gly	005					076	,				0,00	
			val 900	١				905	•				910		
		016	Glu 5				920	}				723	,		
	030	١	Ser			939	5				940	,			
0.45	-		ı Trp		950	)				700	,				200
945 Let	Phe	: Ile	e Lev	Leu 965	Lev	. Phe	e Lei	ı Ala	a Ile 970	Phe	e Ile	Э Туг	Ala	Phe 975	Pro
Ası	туг	: Ala	a Ala 980	Met	. Lys	Le	ı Vai	98!	s Pro	) Phe	e Sei	•			

#### INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/19395

A. CLASSIFICATION OF SUBJECT MATTER										
IPC(7) :C12N 15/11, 15/00; C07K 16/00										
US CL: 536/23.1, 435/440, 530/387.1  According to International Patent Classification (IPC) or to both national classification and IPC										
B. FIELDS SEARCHED										
	ocumentation searched (classification system followed	by classification symbols)								
U.S. :	536/23.1, 435/440, 530/387.1	•								
Documentat	ion searched other than minimum documentation to the	extent that such documents are included	in the fields searched							
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)										
	CAPLUS, EMBASE, ESBIOBASE, LIFESCI, MEDLI rms: dysferlin, lgmd2b	NE, SCISEARCH, TOXLIT	,							
Scarch 16	ims. dysteria, igaidzb									
C. DOCUMENTS CONSIDERED TO BE RELEVANT										
Category*	Citation of document, with indication, where app	propriate, of the relevant passages	Relevant to claim No.							
$ \mathbf{x} $	WEILER et al. Limb-girdle muscu	lar dystrophy and Myoshi	32,35							
	Myopathy in an aboriginal Canadian kindred map to LGMD2B and									
	segregate with the same haplotype. A	T T								
	Genetics. October 1996, Vol.59, pages 872-878, especially page									
	873.									
]	ROENIG	of all Districts Management	22 22 26							
X	KOENIG et al. Complete cloning	3	32-33,36							
	Dystrophy (DMD) cDNA and preliminary genomic organization of									
İ	the DMD gene in normal and affected individuals. Cell. 31 July									
	1987, Vol. 50, pages 509-517, especially pages 511-513.									
	<b>i</b>									
X Further documents are listed in the continuation of Box C. See patent family annex.										
* Special categories of cited documents:  "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand										
	*A* document defining the general state of the art which is not considered the principle or theory underlying the invention to be of particular relevance									
"E" ca	considered novel or cannot be considered to involve an inventive step									
"L" document which may throw doubts on priority claim(s) or which is when the document is taken alone gird to establish the publication date of another citation or other										
special reason (as specified)  "Y"  document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is										
	document referring to an oral disclosure, use, exhibition or other combined with one or more other such documents, such combination being obvious to a person skilled in the art									
	P document published prior to the international filing date but later than ege document member of the same patent family the priority date claimed									
Date of the actual completion of the international search  Date of mailing of the international search report										
17 NOVEMBER 1999 13 JAN 2000										
Name and mailing address of the ISA/US  Authorized officer										
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231  Authorized officer Stephen Siu										
_	Washington, D.C. 20231 Facsimile No. (703) 305-4242 Telephone No. (703) 308-0196									
I I desimile i	, , , , , , , , , , , , , , , , , , ,	, priorie :								

Form PCT/ISA/210 (second sheet)(July 1992)\*

## INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/19395

C-4	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Category*	Citation of document, with indication, where appropriate, of the relevant passages	
X,P	Database GenCore version 4.5, Compugen Ltd., No. AI128455, 'NCI-CGAP, National Cancer Institute, Cancer Genome Anatomy	1,6,12
Y,P	Project (CGAP), Tumor Gene Index', Unpublished, 27 October 1998	7,14,16
x	Database GenCore version 4.5, Compugen Ltd., No. R41062,	1, 6, 11-12
Y	WAYE, M.M.Y. et al. 'Gene expression of adult human heart as revealed by random sequencing of cDNA library,' Miami Winter Biotechnol. Symp. Proc. 6,90, 16 May 16, 1995.	7, 14
x	Database GenCore version 4.5, Compugen Ltd., No. AA718275, Marra et al, 'The WashU-HHMI Mouse EST Project', Unpublished,	1, 6, 11-12
Y	29 December 1997.	7, 14
Y	BASHIR et al. Genetic and physical mapping at the limb-girdle muscular dystrophy locus (LGMD2B) on chromosome 2p. Genomics. April 1996, Vol.33, pages 46-52, especially page 47.	32,36
Y	MOREIRA et al. The seventh form of autosomal recessive limb- girdle muscular dystrophy is mapped to 17q11-12. American Journal of Human Genetics. July 1997, Vol. 61, pages 151-159, entire document.	32, 35
Y	Database GenCore version 4.5, Compugen Ltd., No. R76778, HILLIER et al., 'The WashU-Merck EST Project', Unpublished, 06 June 1995.	7, 14
A,E	AHLBERG et al. Genetic Linkage of Welander Distal Myopathy to chromosome 2p13. Annals of Neurology. September 1999, Vol. 46, No.3, pages 399-404, especially page 400.	37, 39
A,E	BITTNER et al. Dysferlin deletion in SJL mice (SJL-Dysf) defines a natural model for limb girdle muscular dystrophy 2B. Nature Genetics. October 1999, Vol. 23, pages 141-142, especially page 141.	40
A,P	BASHIR et al. A gene related to Caenorhabditis elegans spermatogenesis factor fer-1 is mutated in limb-girdle muscular dystrophy type 2B. Nature Genetics. September 1998, Vol 20, pages 37-42.	1-53
A,E	Matsuda et al. Dysferlin is a surface membrane-associated protein that is absent in Miyoshi Myopathy. Neurology 22 September 1999, Vol. 53, No. 5, pages 1119-1122, especially pages 1119-1120.	40

Form PCT/ISA/210 (continuation of second sheet)(July 1992)\*

ntemational	application	No.
PCT/US99/	19395	

# INTERNATIONAL SEARCH REPORT

C (Continue	tion). DOCUMENTS CONSIDERED TO BE RELEVANT	
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No
A,P	LIU et al. Dysferlin, a novel skeletal muscle gene, is mutated in Miyoshi Myopathy and limb girdle muscular dystrophy. Nature Genetics. September 1998, Vol. 20, pages 31-36.	1-54
	·	

Form PCT/ISA/210 (continuation of second sheet)(July 1992)\*



# PCT

# WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



#### INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7:

(11) International Publication Number:

WO 00/11157

C12N 15/11, 15/00, C07K 16/00

Al

(43) International Publication Date:

2 March 2000 (02.03.00)

(21) International Application Number:

PCT/US99/19395

(22) International Filing Date:

25 August 1999 (25.08.99)

(30) Priority Data:

60/097,927

25 August 1998 (25.08.98)

us

(71) Applicant: THE GENERAL HOSPITAL CORPORATION [US/US]; 55 Fruit Street, Boston, MA 02114 (US).

- (72) Inventors: BROWN, Robert, H.; 16 Oakland Avenue, Needham, MA 02192 (US). LIU, Jing; 1629 Van Horn, Outremont, Quebec H2V IL1 (CA). AOKI, Masashi; Dept. of Neurology, Tohoku University School of Medicine, 1-1, Seiro-Machi, Aoba-ku, Sendai (JP). HO, Meng, F.; Apartment 24, 145 Englewood Avenue, Brighton, MA 02135 (US). MATSUDA-ASADA, Chie; 33 Pond Avenue, Brookline, MA 02445 (US).
- (74) Agent: FRASER, Janis, K.; Fish & Richardson P.C., 225 Franklin Street, Boston, MA 02110-2804 (US).

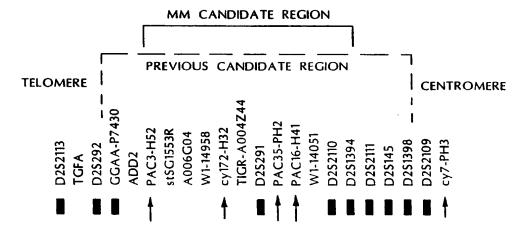
(81) Designated States: CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

#### **Published**

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: DYSFERLIN, A GENE MUTATED IN DISTAL MYOPATHY AND LIMB GIRDLE MUSCULAR DYSTROPHY



#### (57) Abstract

A novel gene and the protein encoded therein, i.e., dysferlin, are disclosed. This gene and its expression products are associated with muscular dystrophy, e.g., Miyoshi myopathy and limb girdle muscular dystrophy 2B.

## FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
ΑT	Austria	FR	France	LU	Luxembourg	SN	Senegal
ΑU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
ΑZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA.	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of Americ
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Vict Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	zw	Zimbabwe
Cı	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
ÐE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

# DYSFERLIN, A GENE MUTATED IN DISTAL MYOPATHY AND LIMB GIRDLE MUSCULAR DYSTROPHY

# RELATED APPLICATION INFORMATION

This application claims priority from provisional application serial no. 60/097,927, filed August 25, 1998.

#### Statement as to Federally Sponsored Research

The work described herein was supported in part by 10 NIH grants 5P01AG12992, 5R01N834913A, and 5P01NS31248.

The Federal Government therefore may have certain rights in the invention.

#### Background of the Invention

The invention relates to genes involved in the 15 onset of muscular dystrophy.

Muscular dystrophies constitute a heterogeneous group of disorders. Most are characterized by weakness and atrophy of the proximal muscles, although in rare myopathies such as "Miyoshi myopathy" symptoms may first 20 arise in distal muscles. Of the various hereditary types of muscular dystrophy, several are caused by mutations or deletions in genes encoding individual components of the dystrophin-associated protein (DAP) complex. It is this DAP complex that links the cytoskeletal protein 25 dystrophin to the extracellular matrix protein, laminin-2.

Muscular dystrophies may be classified according to the gene mutations that are associated with specific clinical syndromes. For example, mutations in the gene encoding the cytoskeletal protein dystrophin result in either Duchenne's Muscular Dystrophy or Becker's Muscular Dystrophy, whereas mutations in the gene encoding the extracellular matrix protein merosin produce Congenital

5

- 2 -

Muscular Dystrophy. Muscular dystrophies with an autosomal recessive mode of inheritance include "Miyoshi myopathy" and the several limb-girdle muscular dystrophies (LGMD2). Of the limb-girdle muscular dystrophies, the deficiencies resulting in LGMD2C, D, E, and F result from mutations in genes encoding the membrane-associated sarcoglycan components of the DAP complex.

# Summary of the Invention

A novel protein, designated dysferlin, is 10 identified and characterized. The dysferlin gene is normally expressed in skeletal muscle cells and is selectively mutated in several families with the hereditary muscular dystrophies, e.g., Miyoshi myopathy 15 (MM) and limb girdle muscular dystrophy-2B (LGMD2B). These characteristics of dysferlin render it a candidate disease gene for both MM and LGMD2B. An additional novel protein, brain-specific dysferlin, has also been identified. Defects in brain-specific dysferlin may 20 predispose to selected disorders of the central nervous Moreover, the expression of brain-specific dysferlin may be important as a marker for normal neural development (e.g., in vivo or in neural cells in culture). Manipulation of levels of expression of brain-25 specific dysferlin, and of the type of expressed brainspecific dysferlin is of use for analyzing the function of brain-specific dysferlin and related dysferlinassociated molecules.

The invention features an isolated DNA which
30 includes a nucleotide sequence hybridizing under
stringent hybridization conditions to a strand of SEQ ID
NO:3 or SEQ ID NO:117.

- 3 -

The invention also features an isolated DNA including a nucleotide sequence selected from SEQ ID NOs:4-12.

Also within the invention is an isolated DNA comprising a nucleotide sequence selected from the group consisting of SEQ ID NOs:22-30.

Also within the invention is a single stranded oligonucleotide of 14-50 nucleotides in length having a nucleotide sequence identical to a portion of a strand of 10 SEQ ID NO:3.

Also within the invention is a pair of PCR primers consisting of:

- (a) a first single stranded oligonucleotide consisting of 14-50 contiguous nucleotides of the sense 15 strand of SEQ ID NO:117; and
- (b) a second single stranded oligonucleotide consisting of 14-50 contiguous nucleotides of the antisense strand of SEQ ID NO:117, wherein the sequence of at least one of the oligonucleotides is identical to a 20 portion of a strand of SEQ ID NO:3, and the first oligonucleotide is not complementary to the second oligonucleotide.

Also within the invention is a pair of single stranded oligonucleotides selected from of SEQ ID NO: 130-231, SEQ ID NO:110, and SEQ ID NO:112.

Also within the invention is an isolated DNA including a nucleotide sequence that encodes a protein that shares at least 70% sequence identity with SEQ ID NO:2, or a complement of the nucleotide sequence.

Also within the invention is an isolated DNA including a nucleotide sequence which hybridizes under stringent hybridization conditions to a strand of a nucleic acid, the nucleic acid having a sequence selected from SEQ ID NOs:31-79 and 90-101.

Also within the invention is a single stranded oligonucleotide of 14-50 nucleotides in length having a nucleotide sequence which is identical to a portion of a strand of a nucleic acid selected from SEQ ID NOs:31-79 and 90-100.

Also within the invention is a pair of PCR primers consisting of:

- (a) a first single stranded oligonucleotide consisting of 14-50 contiguous nucleotides of the sensestrand of a nucleic acid selected from SEQ ID NOs:31-85; and
- (b) a second single stranded oligonucleotide consisting of 14-50 contiguous nucleotides of the antisense strand of a nucleic acid selected from SEQ ID NOs:31-85, wherein the sequence of at least one of the oligonucleotides includes a sequence identical to a portion of a strand of a nucleic acid selected from SEQ ID NOs: 31-79 and 90-100, and the first oligonucleotide is not complementary to the second oligonucleotide.
- Also within the invention is a pair of single stranded oligonucleotides selected from SEQ ID NOs 101-116, SEQ ID NOs 184-185, SEQ ID NOs 188-191, SEQ ID NOs 210-213, and SEQ ID NOs 216-217.

Also within the invention is a substantially pure protein that has an amino acid sequence sharing at least 70% sequence identity with SEQ ID NO:2.

Also within the invention is a substantially pure protein the sequence of which includes amino acid residues 1-500, 501-1000, 1001-1500, or 1501-2080 of SEQ 30 ID NO:2.

Also within the invention is a substantially pure protein including the amino acid sequence of SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, or SEQ ID NO:89.

In another aspect, the invention features a 35 transgenic non-human mammal having a transgene disrupting

PCT/US99/19395 WO 00/11157

- 5 **-**

or interfering with the expression of a dysferlin gene, the transgene being chromosomally integrated into the germ cells of the animal.

Another embodiment of the invention features a 5 method of decreasing the symptoms of muscular dystrophy in a mammal by introducing into a cell of the mammal (e.g., a muscle cell or a muscle precursor cell) an isolated DNA which hybridizes under stringent hybridization conditions to a strand of SEQ ID NO:3.

Another aspect of the invention provides a method for identifying a patient, a fetus, or a pre-embryo at risk for having a dysferlin-related disorder by (a) providing a sample of genomic DNA from the patient, fetus, or pre-embryo; and (b) determining whether the 15 sample contains a mutation in a dysferlin gene.

In another aspect, the invention provides a method for identifying a patient, a fetus, or a pre-embryo at risk for having a dysferlin-related disorder by (a) providing a sample including dysferlin mRNA from the 20 patient, fetus, or pre-embryo; and (b) determining whether the dysferlin mRNA contains a mutation.

Methods of identifying mutations in a dysferlin sequence are useful for predicting (e.g., predicting whether an individual is at risk for developing a 25 dysferlin-related disorder) or diagnosing disorders associated with dysferlin, e.g., MM and LGMD2B. methods can also be used to determine if an individual, fetus, or a pre-embryo is a carrier of a dysferlin mutation, for example in screening procedures. Methods 30 which distinguish between different dysferlin alleles (e.g., a mutant dysferlin allele and a normal dysferlin allele) can be used to determine carrier status.

The invention also features an isolated nucleic acid comprising a nucleotide sequence which hybridizes 35 under stringent hybridization conditions to nucleic acids

10

3284-3720 of SEQ ID NO:232, or the complement of the nucleotide sequence. An isolated nucleic acid including a nucleotide sequence identical to the sequence of nucleotides 3284-3720 of SEQ ID NO:232, or a complement of the nucleotide sequence is also a feature of the invention. The isolated nucleic acid can include the entire sequence of SEQ ID NO:232 or the complement of SEQ ID NO:232.

Another aspect of the invention features an isolated polypeptide that includes: a) at least 15 contiguous amino acids of the polypeptide comprising amino acids 1-24 of SEQ ID NO:233, b) a naturally occuring allelic variant of a polypeptide comprising amino acids 1-24 of SEQ ID NO:233, or c) an amino acids 1-24 of SEQ ID NO:233, or c) an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes under stringent conditions to nucleotides 3284-3720 of SEQ ID NO:232. The polypeptide of this aspect can include the entire sequence of SEQ ID NO:233.

Also included in the invention is a vector comprising the nucleic acid of claim 44 and a cell that contains the vector. Another aspect of the invention features a method of making a polypeptide by culturing the cell which contains the vector.

The invention also features an antibody which specifically binds to a polypeptide of such as those described above. The antibody can bind to a polypeptide selected from amino acids 253-403 of SEQ ID NO:233, amino acids 624-865 of SEQ ID NO:233, and amino acids 1664-1786 of SEQ ID NO:233. Antibodies of the invention can be monclonal or polyclonal antibodies.

An "isolated DNA" is DNA which has a naturally occurring sequence corresponding to part or all of a given gene but is free of the two genes that normally flank the given gene in the genome of the organism in

- 7 -

which the given gene naturally occurs. The term therefore includes a recombinant DNA incorporated into a vector, into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote. It also includes a separate molecule such as a cDNA, a genomic fragment, a fragment produced by polymerase chain reaction (PCR), or a restriction fragment, as well as a recombinant nucleotide sequence that is part of a hybrid gene, i.e., a gene encoding a fusion protein. The term excludes intact chromosomes and large genomic segments containing multiple genes contained in vectors or constructs such as cosmids, yeast artificial chromosomes (YACs), and P1-derived artificial chromosome (PAC) contigs.

A "noncoding sequence" is a sequence which corresponds to part or all of an intron of a gene, or to a sequence which is 5' or 3' to a coding sequence and so is not normally translated.

An expression control sequence is "operably
linked" to a coding sequence when it is within the same
nucleic acid and can control expression of the coding
sequence.

A "protein" or "polypeptide" is any chain of amino acids linked by peptide bonds, regardless of length or post-translational modification, e.g., glycosylation or phosphorylation.

As used herein, the term "percent sequence identity" means the percentage of identical subunits at corresponding positions in two sequences when the two sequences are aligned to maximize subunit matching, i.e., taking into account gaps and insertions. For purposes of the present invention, percent sequence identity between two polypeptides is to be determined using the Gap program and the default parameters as specified therein.

35 The Gap program is part of the Sequence Analysis Software

15

Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705.

The algorithm of Myers and Miller, CABIOS (1989)

5 can also be used to determine whether two sequences are similar or identical. Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a

10 PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used.

As used herein, the term "stringent hybridization conditions" means the following DNA hybridization and wash conditions: hybridization at 60°C in the presence of 6 x SSC, 0.5% SDS, 5 x Denhardt's Reagent, and 100 µg/ml denatured salmon sperm DNA; followed by a first wash at room temperature for 20 minutes in 0.5 x SSC and 0.1% SDS and a second wash at 55°C for 30 minutes in 0.2 x SSC and 0.1% SDS.

A "substantially pure protein" is a protein 20 separated from components that naturally accompany it. The protein is considered to be substantially pure when it is at least 60%, by dry weight, free from the proteins and other naturally-occurring organic molecules with 25 which it is naturally associated. Preferably, the purity of the preparation is at least 75%, more preferably at least 90%, and most preferably at least 99%, by weight. A substantially pure dysferlin protein can be obtained, for example, by extraction from a natural source, by 30 expression of a recombinant nucleic acid encoding a dysferlin polypeptide, or by chemical synthesis. Purity can be measured by any appropriate method, e.g., column chromatography, polyacrylamide gel electrophoresis, or HPLC analysis. A chemically synthesized protein or a 35 recombinant protein produced in a cell type other than

- 9 -

the cell type in which it naturally occurs is, by definition, substantially free from components that naturally accompany it. Accordingly, substantially pure proteins include those having sequences derived from 5 eukaryotic organisms but which have been recombinantly produced in E. coli or other prokaryotes.

An antibody that "specifically binds" to an antigen is an antibody that recognizes and binds to the antigen, e.g., a dysferlin polypeptide, but which does 10 not substantially recognize and bind to other molecules in a sample (e.g., a biological sample) which naturally includes the antigen, e.g., a dysferlin polypeptide. antibody that "specifically binds" to dysferlin is sufficient to detect a dysferlin polypeptide in a 15 biological sample using one or more standard immunological techniques (for example, Western blotting or immunoprecipitation).

A "transgene" is any piece of DNA, other than an intact chromosome, which is inserted by artifice into a 20 cell, and becomes part of the genome of the organism which develops from that cell. Such a transgene may include a gene which is partly or entirely heterologous (i.e., foreign) to the host organism, or may represent a gene homologous to an endogenous gene of the organism.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Methods and materials similar or equivalent to those described herein can be 30 used in the practice or testing of the present invention. The present materials, methods, and examples are illustrative only and not intended to be limiting. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference 35 in their entirety. In case of conflict, the present

25

- 10 -

specification, including definitions, will control. All the sequences disclosed in the sequence listing are meant to be double-stranded except the sequences of oligonucleotides.

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

#### Brief Description of the Drawings

Fig. 1A is a physical map of the MM locus. Arrows indicate the five new polymorphic markers and filled, vertical rectangular boxes indicate the previously known polymorphic markers. The five ESTs that are expressed in skeletal muscle are highlighted in bold. Detailed information on the minimal tiling path of the PAC contig spanning the MM/LGMD2B region is provided in Liu et al., 1998, Genomics 49:23-29. The minimal candidate MM region is designated by the solid bracket (top) and compared to the previous candidate region (dashed bracket). TGFA and ADD2 are transforming growth factor alpha and β-adducin 20 2.

Fig. 1B is a representation of the dysferlin cDNA clones. The probes used in the three successive screens are shown in bold (130347, cDNA10, A27-F2R2). The two most 5' cDNA clones are also shown (B22, B33). The 6.9

25 kb cDNA for dysferlin (SEQ ID NO:1) is illustrated at the bottom with start and stop codons as shown.

Fig. 1C is a representation of the predicted dysferlin protein. The locations of four C2 domains (SEQ ID NOs: 86-89) are indicated by stippled boxes,

while the putative transmembrane region is hatched.

Vertical lines above the cDNA denote the positions of the mutations in Table 2; the associated labels indicate the phenotypes (MM - Miyoshi myopathy; LGMD - limb girdle

PCT/US99/19395 WO 00/11157

- 11 -

muscular dystrophy; DMAT - distal myopathy with anterior tibial onset).

Fig. 2 is the sequence of the predicted 2,080 amino acids of dysferlin (SEQ ID NO:2). The predicted 5 membrane spanning residues are in bold at the carboxy terminus (residues 2047-2063). Partial C2 domains are underlined. Bold, underlined sequences are putative nuclear targeting residues. Possible membrane retention sequences are enclosed within a box.

Fig. 3 is a comparison of the Kyle-Doolittle hydrophobicity plots of the dysferlin protein and fer-1. On the Y-axis, increasing positivity corresponds to increasing hydrophobicity. Both proteins have a single, highly hydrophobic stretch at the carboxy terminal end 15 (arrow). Both share regions of relative hydrophilicity approximately at residue 1,000 (arrowhead).

Fig. 4 is a SSCP analysis of a representative pedigree with dysferlin mutations. Each member of the pedigree is illustrated above the corresponding SSCP 20 analysis. For each affected individual (solid symbols) shifts are evident in alleles 1 and 2, corresponding respectively to exons 36 and 54. As indicated, the allele 1 and 2 variants are transmitted respectively from the mother and the father. The two affected daughters in 25 this pedigree have the limb girdle muscular dystrophy (LGMD) phenotype while their affected brother has a pattern of weakness suggestive of Miyoshi myopathy (MM).

Fig. 5 is a representation of the genomic structure of dysferlin. The 55 exons of the dysferlin 30 gene and their corresponding SEQ ID NOs are indicated below the 6911 bp cDNA (solid line). The cDNA sequences corresponding to SEQ ID NO:1 and SEQ ID NO:3 are shown relative to the 6911 bp cDNA.

Figs. 6A-B are the cDNA sequence of brain-specific 35 dysferlin (SEQ ID NO:232) and the predicted amino acid

10

sequence (in single-letter code) of brain-specific dysferlin (SEQ ID NO:233).

#### Detailed Description

The Miyoshi myopathy (MM) locus maps to human 5 chromosome 2p12-14 between the genetic markers D2S292 and D2S286 (Bejaoui et al., 1995, Neurology 45:768-72). Further refined genetic mapping in MM families placed the MM locus between markers GGAA-P7430 and D2S2109 (Bejaoui et al., 1998, Neurogenetics 1:189-96). Independent 10 investigation has localized the limb-girdle muscular dystrophy (LGMD-2B) to the same genetic interval (Bashir et al., 1994, Hum. Molec. Genetics 3:455-57; Bashir et al., 1996, Genomics 33:46-52; Passos-Bueno et al., 1995, Genomics 27:192-95). Furthermore, two large, inbred 15 kindreds have been described whose members include both MM and LGMD2B patients (Weiler et al., 1996, Am. J. Hum. Genet. 59:872-78; Illarioshkin et al., 1997, Genomics 42:345-48). In these familial studies, the disease gene(s) for both MM and LGMD2B mapped to essentially the 20 same genetic interval. Moreover, in both pedigrees, individuals with MM or LGMD2B phenotypes share the same haplotypes. This raises the intriguing possibility that the two diseases may arise from the same gene defect and that a particular disease phenotype is the result of 25 modification by additional factors.

A 3-Mb PAC contig spanning the entire MM/LGMD2B candidate region was recently constructed to facilitate the cloning of the MM/LGMD2B gene(s) (Liu et al., 1998, Genomics 49:23-29). This high resolution PAC contig resolved the discrepancies of the order of markers in previous studies (Bejaoui et al., 1998, Neurogenetics 1:189-96; Bashir et al., 1996, Genomics 33:46-52; Hudson et al., 1995, Science 270:1945-54). The physical size of the PAC contig also indicated that the previous minimal

- 13 -

size estimation based on YAC mapping data was significantly underestimated.

#### Identification of Repeat Sequences and Repeat Typing

The PAC contig spanning the MM/LGMD2B region (Liu et al., 1998, Genomics 49:23-29) was used as a source for the isolation of new informative markers to narrow the genetic interval of the disease gene(s). DNA from the PAC clones spanning the MM/LGMD2B region was spotted onto Hybond N+™ membrane filters (Amersham, Arlington Heights, IL). The filters were hybridized independently with the following γ-32P (Du Pont, Wilmington, DE) labeled repeat sequences: (1) (CA)<sub>15</sub>; (2) pool of (ATT)<sub>10</sub>, (GATA)<sub>8</sub> and (GGAA)<sub>8</sub>; (3) pool of (GAAT)<sub>8</sub>, (GGAT)<sub>8</sub> and (GTAT)<sub>8</sub>; and (4) pool of (AAG)<sub>10</sub> and (ATC)<sub>10</sub>. Hybridization and washing of the filters were carried out at 55°C following standard protocols (Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual (2nd Edition), Cold Spring Harbor Press, N.Y.).

Miniprep DNAs of PAC clones containing repeat 20 sequences were digested with restriction enzymes HindIII and PstI and ligated into pBluescript II (KS+) vector which is (Stratagene, La Jolla, CA) digested with the same enzymes. Filters of the PAC subclones were hybridized to the  $\gamma$ -32P labeled repeats that detected the 25 respective PACs. For clones with an insert size greater than 1 kb the repeat sequences of which could not be identified by a single round of sequencing, the inserts were further subcloned by digestion with HaeIII and ligation in EcoRV-digested pZero-2.1 vector (Invitrogen, 30 Inc., Carlsbad, CA). Miniprep DNAs of the positive subclones were subjected to manual dideoxy sequencing with Sequenase™ enzyme (US Biochemicals, Inc., Cleveland, OH). Primer pairs for amplifying the repeat sequences were selected using the computer program Oligo (Version

- 14 -

4.0, National Biosciences, Inc., Plymouth, MN). Primer sequences are shown in Table 1.

- 15 -

TABLE 1

			01		-1	2
	Het.	0.82	0.72	0.30	0.41	0.32
Region	No. of alleles	10	٢	<b>.</b>	44	4
MM/LGMD2B I	Size in PAC (bp)	138	199	161	280	211
Mapped to the	Annealing Im (°C)	7.50	20	26	80	<b>9</b>
New Polymorphic Markers Mapped to the MM/LGMD2B Region	Primers (5' to 3')	GATCTAACCCTGCTGCTCACC (SEQ ID NO:120) CTGGTGTTGCAGAGCGCTG (SEQ ID NO:121)	CCTCTTCTGCTGTCTTCAG (SEQ ID NO:122) TGTGTCTGGTTCCACCTTCGT (SEQ ID NO:123)	TCCAAATAGAAATGCCTGAAC (SEQ ID NO:124) AGGTATCACCTCCAAGTGTTG (SEQ ID NO:125)	TACCAGCTTCAGAGCTCCCTG (SEQ ID NO:126) TTGATCAGGTGCTCTTGG (SEQ ID NO:127)	GGAGAATTGCTTGAACCCAG (SEQ ID NO:128) TGGCTAATGATGTTGAACATTT (SEQ ID NO:129)
	Repeat	CA	CCAT	CAT	Complex	AAGG
	Marker	PAC3-H52	Су172-н32³	PAC35-PH2	PAC16-H41	Су7 - РНЗ

Observed in 50 unrelated caucasians.

Heterozygosity index. Located within intron 2 of the dysferlin gene.

All oligonucleotides were synthesized by Integrated DNA Technologies, Inc. (Coralville, IA). PCR typing of the repeat markers followed previously described protocols (Bejaoui et al., 1995, Neurology 45:768-772).

Identification of Repeat Markers and Haplotype Analysis After hybridization with labeled repeat oligos, 17 different groups of overlapping PACs were identified that contained repeat sequences. Some groups contained 5 previously identified repeat markers. For example, five groups of PACs were positively identified by a pool of repeat probes including (ATT) 10, (GATA) 8, and (GGAA) 8. Of these, three groups contained known markers GGAA-P7430 (GGAA repeat), D2S1394 (GATA repeat) and D2S1398 (GGAA 10 repeat) (Hudson et al., 1992, Nature 13:622-29; Gastier et al., 1995, Hum. Molecular Genetics 4:1829-36). attempt was made to isolate new repeat markers from these PACs and they were not further analyzed. Similarly, seven groups of PACs that contained known CA repeat 15 markerswere excluded. Seven groups of PACs that contained unidentified repeats were retained for further analysis. For each group, the PAC containing the smallest insert was selected for subcloning. Subclones were re-screened and positive clones were sequenced to

20 identify repeats. In total, seven new repeat sequences were identified within the MM/LGMD2B PAC contig. Of these, five are polymorphic within the population that was tested. The information for these five markers is summarized in Table 1. Based on the PAC contig

25 constructed previously across the MM candidate locus (Liu et al., 1998, *Genomics* 48:23-29), the five new markers and ten previously published polymorphic markers were placed in an unambiguous order (Fig. 1).

These markers were analyzed in a large,

30 consanguineous MM family (Bejaoui et al., 1995, Neurology
45: 768-72; Bejaoui et al., 1998, Neurogenetics 1:18996). Because MM is a recessive condition, the locus can
be defined by identifying regions of the genome that show
homozygosity in affected individuals. Conversely,
35 because of the high penetrance of this adult-onset

- 17 -

condition, unaffected adult individuals are not expected to be homozygous by descent across the region. Analysis of haplotype homozygosity in this pedigree indicates that the disease gene lies between markers D2S2111 and PAC3-H52. Based on the PAC mapping data, the physical distance for this interval is approximately 2.0 Mb. No recombination events were detected between four informative markers (markers cy172-H32 to PAC16-H41) and the disease locus in family MM-21 (Fig. 1A).

#### 10 Identification of Five Muscle-Expressed ESTs

Twenty-two ESTs and two genes (transforming growth factor alpha [TGFα] and beta-adducin [ADD2]) were previously mapped to the MM/LGMD2B PAC contig (Fig. 1A) (Liu et al., 1998, Genomics 48:23-29). Two μl

15 (approximately 0.1 ng/μl) of Marathon-ready™ skeletal muscle cDNA (Clontech, Palo Alto, CA) were used as template in a 10 μl PCR reaction for analysis of muscle expression of ESTs. The PCR conditions were the same as for the PCR typing of repeat markers. PCR analysis of skeletal muscle cDNA indicated that five of these ESTs (A006G04, stSG1553R, WI-14958, TIGR-A004Z44 and WI-14051) map within the minimal genetic MM interval of MM and are expressed in skeletal muscle.

Probes were selected corresponding to each of
these five ESTs for Northern blot analysis. cDNA clones
(130347, 48106, 172575, 184080, and 510138) corresponding
to the five ESTs that are expressed in muscle
(respectively TIGR-A004Z44, WI-14051, WI-14958, stSG1553R
and A006G04) were selected from the UniGene database
(http:/www.ncbi.nlm.nih.gov/UniGene/) and obtained from
Genome Systems, Inc. (St. Louis, MO). The cDNA probes
were first used to screen the MM/LGMD2B PAC filters to
confirm that they mapped to the expected position in the
MM/LGMD2B contig.

A Northern blot (Clontech) of multiple human tissues was sequentially hybridized to the five cDNA probes and a control  $\beta$ -actin cDNA at 65°C following standard hybridization and washing protocols (Sambrook et al., supra). Between hybridizations, probes were removed by boiling the blot at 95-100°C for 4-10 min with 0.5% SDS. The blot was then re-exposed for 24 h to confirm the absence of previous hybridization signals before proceeding with the next round of hybridization.

The tissue distribution, intensity of the signals and size of transcripts detected by the five cDNA probes varied. Probes corresponding to ESTs stSG1553R, TIGR-A004Z44 and WI-14958 detected strong signals in skeletal muscle. In addition, the cDNA corresponding to TIGR-

15 A004Z44 detected a 3.6-3.8 kb brain-specific transcript instead of the 8.5 kb message that was present in other tissues. It is likely that these five ESTs correspond to different genes since the corresponding cDNA probes used for Northern analysis derive from the 3' end of messages,

20 map to different positions in the MM/LGMD2B contig (Fig. 1A), and differ in their expression patterns.

Current database analysis suggests that three of these ESTs (stSG1553R, WI-14958 and WI-14051) do not match any known proteins (Schuler et al., 1996, Science 274:540-46). A006G04 has weak homology with a protein sequence of unknown function that derives from C. elegans. TIGR-A004Z44 has homology only to subdomains present within protein kinase C. Because the five genes corresponding to the ESTs are expressed in skeletal 30 muscle and map within the minimal genetic interval of the MM/LGMD2B gene(s), they are candidate MM/LGMD2B gene(s).

- 19 -

#### Cloning of Dysferlin cDNA

EST TIGR-A004Z44 gave a particularly strong skeletal muscle signal on the Northern blot. Moreover, it is bracketed by genetic markers that show no recombination with the disease phenotype in family MM-21 (Fig. 1). The corresponding transcript was therefore cloned and analyzed as a candidate MM gene. From the Unigene database, a cDNA IMAGE clone (130347, 979 bp) was identified that contained the 483 bp EST TIGR-A004Z44.

Approximately 1 x 10<sup>6</sup> recombinant clones of a λgt11 human skeletal muscle cDNA library (Clontech) were plated and screened following standard techniques (Sambrook et al., supra). The initial library screening was performed using the insert released from the clone 130347 that

15 contains EST TIGR-A0044Z44, corresponding to the 3' end of the gene. Positive phages were plaque purified and phage DNA was isolated according to standard procedures (Sambrook et al., supra). The inserts of the positive clones were released by EcoRI digestion of phage DNA and subsequently subcloned into the EcoRI site of pBluescript II (KS+) vector (Stratagene).

Fifty cDNA clones were identified when a human skeletal muscle cDNA library was screened with the 130347 cDNA. Clone cDNA10 with the largest insert (~6.5 kb)

25 (Fig. 1B) was digested independently with BamHI and PstI and further subcloned into pBluescript vector. Miniprep DNA of cDNA clones and subclones of cDNA10 was prepared using the Qiagen plasmid Miniprep kit (Valencia, CA). Sequencing was carried out from both ends of each clone using the SequiTherm EXCEL<sup>TM</sup> long-read DNA sequencing kit (Epicenter, Madison, WI), fluorescent-labeled M13 forward and reverse primers, and a LI-COR sequencer (Lincoln, NE). Assembly of cDNA contigs and sequence analysis were performed using Sequencher software (Gene Codes

35 Corporation, Inc., Ann Arbor, MI).

- 20 -

Two additional screens, first with the insert of cDNA10 and then a 683 bp PCR product (A27-F2R2) amplified from the 5' end of the cDNA contig, identified 87 additional cDNA clones. Clones B22 and B33 extended the 5' end by 94 and

20 bp, respectively. The compiled sequence allowed for the generation of a sequence of 6.9 kb (SEQ ID NO:1) (with 10-fold average coverage).

Although the 5' end of the gene has not been further extended to the 8.5 kb predicted by Northern analysis, an open reading frame (ORF) of 6,243 bp has been identified within this 6.9 kb sequence. This ORF is preceded by an in-frame stop codon and begins with the sequence cgcaagcATGCTG (SEQ ID NO:118); five of the first seven bp are consistent with the Kozak consensus sequence for a start codon (Kozak, 1989, Nucl. Acids Res. 15:8125-33; Kozak, 1989, J. Cell. Biol. 108:229-41). An alternate start codon, in the same frame, +75 bp downstream, appears less likely as a start site GAGACGATCGGG (SEQ ID NO:119). Thus, the entire coding region of this candidate gene is believed to have been identified, as represented by the 6.9 kb sequence contig.

# Isolation of the Brain-Specific Dysferlin Isoform Identification of the brain-specific isoform of dysferlin

A brain-specific isoform of dysferlin was identified using Northern blot analysis of poly(A+)RNA derived from multiple human adult tissues probed with radiolabeled full-length dysferlin cDNA subclones. A prominent 7.2 kb transcript was detected on Northern blots in skeletal muscle, heart, placenta, lung, and kidney, while a distinct but equally prominent 3.6 kb-3.8 kb transcript was identified exclusively in the brain. Using long exposures, a faint 7.2 kb mRNA was also detected in the

25

- 21 -

brain. This finding suggested that the shorter brain isoform was likely to be a tissue-specific splice variant of the dysferlin gene. To test this hypothesis, a human brain cDNA library (Stratagene) was screened for the dysferlin brain isoform.

Cloning of the brain-specific dysferlin isoform

To identify probes that hybridize to the brainspecific dysferlin sequence and so could be used for
library screening, fragments of the full-length dysferlin

CDNA clone (derived from a skeletal muscle cDNA library)
were generated using restriction enzymes. The fragments
were about 1 kb in length and were analyzed by
hybridization to a Northern blot that included brain RNA.
Sequences suitable for library screening were those that
hybridized to the 3.6-3.8 kb brain-specific transcript.
A region of the 3' end of the dysferlin cDNA sequence
that is approximately 3 kb in length was identified as
hybridizing to brain mRNA. DNA containing sequence from
this region was used as a probe for hybridization

screening of a human brain cDNA library (Stratagene).

The human brain cDNA library was plated out and screened using standard procedures. Of the approximately 720,000 plaques screened, 63 primary positive clones were identified. Of these, 20 clones were selected for further analysis involving standard methods of hybridization, restriction enzyme mapping, and sequencing. The primary positive clones shared regions of overlap with each other.

Sequencing of positive clones, provided 3671
30 nucleotides of the brain-specific dysferlin sequence (SEQ ID NO:232; Figure 6A-B). The identified sequence corresponds closely to the size of the brain-specific dysferlin transcript detected on Northern blots. With the exception of the 5' region of the sequence, the

brain-specific sequence is identical to about 3.1 kb of the dysferlin sequence (from nucleotide 3722 to 6904 of the dysferlin sequence). In the dysferlin gene, position 3722 corresponds to the start of exon 32. This finding is consistent with the hypothesis that the brain isoform is a splice-variant of the dysferlin gene. At the 5' end of the brain isoform, 489 nucleotides are unique to brain-specific dysferlin. The amino acid sequence encoded by the brain dysferlin nucleic acid sequence (SEQ ID NO:233; Figure 6) contains a unique sequence with an initiation codon within a Kozak consensus sequence. The nucleic acid sequence unique to brain-specific dysferlin encodes a novel 24 amino acid sequence.

#### Identification of Mutations in Miyoshi Myopathy

15 Two strategies were used to determine whether this 6.9 kb cDNA (SEQ ID NO:1) is mutated in MM. First, the genomic organization of the corresponding gene was determined and the adjoining intronic sequence at each of the 55 exons which make up the cDNA was identified. 20 identify exon-intron boundaries within the gene, PAC DNA was extracted with the standard Qiagen -Mini Prep protocol. Direct sequencing was performed with DNA Sequence System (Promega, Madison, WI) using 32P endlabeled primers (Benes et al., 1997, Biotechniques 23:98-25 100). Exon-intron boundaries were identified as the sites where genomic and cDNA sequences diverged. Second, in patients for whom muscle biopsies were available, RT-PCR was also used to prepare cDNA for the candidate gene from the muscle biopsy specimen.

Single strand conformational polymorphism analysis (SSCP) was used to screen each exon in patients from 12 MM families. Putative mutations identified in this way were confirmed by direct sequencing from genomic DNA using exon-specific intronic primers. Approximately 20

ng of total genomic DNA from immortalized lymphocyte cell lines were used as a template for PCR amplification analysis of each exon using primers (below) located in the adjacent introns. SSCP analysis was performed as 5 previously described (Aoki et al., 1998, Ann. Neurol. 43:645-53). In patients for whom muscle biopsies were available, mRNA was isolated using RNA-STAT-60™ (Tel-Test, Friendswood, TX) and first-strand cDNA was synthesized from 1-2  $\mu$ g total RNA with MMLV reverse 10 transcriptase and random hexamer primers (Life Technologies, Gaithersburg, MD). Three  $\mu$ l of this product were used for PCR amplification. Eight sets of primers were designed for muscle cDNA, and overlapping cDNA fragments suitable for SSCP analysis were amplified. 15 After initial denaturation at 94°C for 2 min, amplification was performed using 30 cycles at 94°C for 30

s, 56°C for 30 s, and 72°C for 60 s. The sequences of polymorphisms detected by SSCP analysis were determined by the dideoxy termination method using the Sequenase kit (US Biochemicals). In some instances, the base pair changes predicted corresponding changes in restriction enzyme recognition sites. Such alterations in restriction sites were verified by digesting the relevant PCR products with the appropriate restriction enzymes.

25 Primer pairs used for SSCP screening and exon sequencing are as follows:

- (1) exon 3, F3261 5'-tctcttctcctagagggccatag-3' (SEQ ID NO: 101) and R326 5'-ctgttcctcccatcgtctcatgg-3' (SEQ ID NO: 102);
- 30 (2) exon 20, F3121 5'-gctcctcccgtgaccctctg-3' (SEQ ID NO: 103) and R3121 5'-gggtcccagccaggagcactg-3' (SEQ ID NO: 104);
- (3) exon 36, F2102 5'-cccctctcaccatctcctgatgtg-3'
  (SEQ ID NO: 105) and R2111 5'-tggcttcaccttccctctacctcgg35 3' (SEQ ID NO: 106);

(4) exon 49, F1081 5'-tcctttggtaggaaatctaggtgg-3' (SEQ ID NO: 107) and R1081 5'-ggaagctggacaggcaagagg-3' (SEQ ID NO: 108); (5) exon 50, F1091 5'-atatactgtgttggaaatcttaatgag-3' 5 (SEQ ID NO: 109) and R1091 5'-gctggcaccacagggaatcgg-3' (SEQ ID NO: 110); (6) exon 51, F1101 5'-ctttqcttccttqcatccttctctq-3' (SEQ ID NO: 111) and R1101 5'-agcccccatgtgcagaatggg-3' (SEQ ID NO: 112); (7) exon 52, F1111 5'-ggcagtgatcgagaaacccgg-3' (SEQ 10 ID NO: 113) and R1111 5'-catgccctccactggggctgg-3' (SEQ ID NO: 114); (8) exon 54, F1141 5'-ggatgcccagttgactccggg-3' (SEQ ID NO: 115) and R1141 5'-ccccaccacagtgtcgtcagg-3' (SEQ ID NO: 15 116); (9) exon 29, F3031 5'-aagtgccaagcaatgagtgaccgg-3' (SEQ ID NO: 184) and R3021 5'-ctcactcccacccaccacctq-3' (SEQ ID NO: 185); (10) exon 31, F2141 5'-gaatctgccataaccagcttcgtg-3' (SEQ 20 ID NO: 188) and R2141 5'-tatcaccccatagaggcctcgaag-3' (SEQ ID NO: 189); (11) exon 32, F2981 5'-cagccactcactctggcacctctg-3' (SEQ ID NO: 190) and R2981 5'-agcccacagtctctgactctcctg-3' (SEQ ID NO: 191); 25 (12) exon 43, F2031 5'-cagccaaaccatatcaacaatg-3' (SEQ ID NO: 210) and R2021 5'-ctggggaggtgagggctctag-3' (SEQ ID NO: 211); (13) exon 44, F2011 5'-gaagtqttttqtctcctcctc-3' (SEQ ID NO: 212) and R2011 5'-gcaggcagccagccccatc-3' (SEQ ID NO: 30 213); (14) exon 46, F1041 5'-ctcgtctatgtcttgtgcttgctc-3' (SEQ ID NO: 216) and R1051 5'-caccatggtttggggtcatgtgg-3' (SEQ ID

NO: 217).

- 25 -

These primers were used in SSCP screening and exon sequencing, and identified eighteen different mutations in fifteen families (Table 2).

BNSDOCID: <WO\_\_\_\_\_0011157A1\_IA>

Мате	Nucleotide Change	Exon	Consequence	Origin	Family name	Allele	Change of restriction site
Mutations 537insA	ins of A at 537	е	Frameshift	Arabic	MM59	Нош	no change
Q605X	<u>C</u> AG to <u>T</u> AG at 2186	20	Stop at 605	French	MM67	Hom	-Pst I, -Fnu 4H I¹
I1298V	ATC to GTC at 4265	36	Amino acid change	Italian	MM, LGMD56	Het	-BamHI, -BStYI; +Ava II
E1883X	<u>G</u> AG to <u>T</u> AG at 5870	4 9	Stop at 1883	English	MM8	Het	no change
H1857R	C <u>A</u> T to C <u>G</u> T at 5943	20	Amino acid change	English	MM50	Het	no change

no change	change	no change	no change	-Fnu4HI	-HinPI, -Fsp I	-Mboll	-ScrFI, -BstNI, +MaeI, +BfaI
ou	ou	ou	ou	-Fn	-Hi	dM-	-SCrF -BstN +MaeI +BfaI
Нош	Нош	Het	Het	Het	Нош	Hom	Нош
DMAT71	MM75	MM58	MM8	MM56	MM10	MM17	MM4 6
Spanish	Spanish	English	English	Italian	Japanese	Japanese	Mexican
Frameshift	Frameshift	Frameshift	5' splice site	Amino acid change	Amino acid change	Frameshift	Stop at 1160
20	20	51	52	54	59	31	3 2
del of G at 5966	del of G at 5966	del of AG at 6071/6072	<b>Ggt to G<u>a</u>t</b> at 6319+1	CGT to IGT at 6497	CGC to CAG at 3510	del of G at 3746	<u>C</u> AG to <u>T</u> AG at 3851
5966de1G	5966delG	6071/6072de 1AG	6319+1G to A	R2042C	R1046H	3746delG	Q1160X
			Ŋ				10

no change	+Dde I	-Bpm I, -BanII + AvaII, +Sau96I	-Mbo II	Ø
Het	Hom	Hom	Het	Italian MM69
MM14	MM12	ММ63	MM73	
Japanese	Japanese	French	Spanish	Frameshift
Frameshift	Stop at 1586	Frameshift	Stop at 1732	of ACCCA at 23 e provide -77
4, 8	43	4.	46	of ACCCA e provide
del of CA at 5122/5123, A to T at 5121	<u>C</u> GA to <u>T</u> GA at 5129	del of G at 5245 and G to C at 5249, or G to C at 5245 and del G at 5249	<u>G</u> AG to <u>T</u> AG at 5567	Del ?Pleas 2573
5122/5123de lCA	R1586X	5245delG	E1732X	2573-77 Hom del ACCCA

Ŋ

<sup>1</sup> MM: Miyoshi myopathy; DMAT: distal myopathy with anterior tibial onset; LGMD: limb girdle muscular dystrophy <sup>2</sup> +: create a new restriction site, -: eliminate an existing restriction site.

Twelve of the eighteen different mutations are predicted to block dysferlin expression, either through nonsense or frameshift changes. Seven of the thirteen samples are homozygous and thus expected to result in complete loss 5 of dysferlin function. For each mutated exon in these patients, at least 50 control DNA samples (100 chromosomes) were screened to determine the frequencies of the sequence variants. When possible, the parents and siblings of affected individuals were also screened to 10 verify that defined mutations were appropriately coinherited with the disease in each pedigree (Fig. 4). In two families (50, 58 in Table 2) heterozygous mutations were identified in one allele (respectively a missense mutation and a 2 bp deletion). Mutations in the other 15 allele are presumed to have not been detected (or in three of the screened MM families) either because the mutant and normal SSCP products are indistinguishable or because the mutation lies outside of coding sequence (i.e., in the promoter or a regulatory region of an 20 intron). The disease-associated mutations did not appear to arise in the population as common polymorphisms.

More mutations can be identified by using appropriate primer pairs to amplify an exon and analyze its sequence. The following primer pairs are useful for 25 exon amplification.

	Exon Code		Primer Sequence
	1	F408	5'-gacccacaagcggcgcctcgg-3'{SEQ ID
	NO: 130}		
		F4101	5'-gaccccggcgagggtggtcgg-3'{SEQ ID
30	NO: 131}		
	2	F4111	5'-tgtctctccattctcccttttgtg-3'{SEQ ID
	NO:132}		
		R4111	5'-aggacactgctgagaaggcacctc-3'{SEQ ID
	NO: 133}		

			- 31 <i>-</i>
	3	F3262	5-agtgccctggtggcacgaagg-3' {SEQ ID
	NO: 134}	R3261	5-cctacctgcaccttcaagccatgg-3' {SEQ ID
	NO: 135}	RSZOI	5 Cocacooguacocoaagecacgg c (c= <b>2</b> co
5	4	F3251	5-cagaagagccagggtgccttagg-3' {SEQ ID
	NO: 136}		
	NO 127	R3251	5-ccttggaccttaacctggcagagg-3' {SEQ ID
	NO: 137}	F3242	5-cgaggccagcgcaccaacctg-3' {SEQ ID
10	NO: 138}		
	_	R3242	5-actgccggccattcttgctggg-3' {SEQ ID
	NO: 139}		
	6	F3231	5-ccaggcctcattagggccctc-3' {SEQ ID
15	NO: 140}	R3231	5-ctgaagaggagcctggggtcag-3' {SEQ ID
	NO: 141}	NS251	5 669aa9a99a9669999669 6 (22 <b>x</b> 22
	7	F3222	5-ctgagatttctgactcttggggtg-3' {SEQ ID
	NO: 142}		,
2.0	NO 142)	R3211	5-aaggttctgccctcatgccccatg-3' {SEQ ID
20	NO: 143}	F3561	5-ctggcctgagggatcagcagg-3' {SEQ ID
	NO: 144}		
		R3561	5-gtgcatacatacagcccacggag-3' {SEQ ID
	NO: 145}		
25	9 NO: 146}	F3551	5-gagctattgggttggccgtgtggg-3' {SEQ ID
	110: 140}	R3552	5-accaacacggagaagtgagaactg-3' {SEQ ID
	NO: 147}		
	10	F3201	5-ccacactttatttaacgctttggcgg-3'{SEQ
30	ID NO: 14	-	,
	NO. 1401	R3201	5-cagaaccaaaatgcaaggatacgg-3' {SEQ ID
	NO: 149}	F3191	5-cttctgattctgggatcaccaaagg-3' {SEQ
	ID NO: 15		J JJJ

- 32 -

	NO 151	F3191	5-ggaccgtaaggaagacccaggg-3' {SEQ ID
	NO: 151}	T2101	
	12	F3181	5-cctgtgctcaggagcgcatgaagg-3'{SEQ ID
	NO: 152}		_
5		R3181	5-gcagaceteceacecaagggeg-3' {SEQ ID
	NO: 153}		
	13	F3171	5-gagacagatgggggacagtcaggg-3' {SEQ ID
	NO: 154}		
		R3171	5-cctcccgagagaaccctcctg-3' {SEQ ID
10	NO: 155}		
	14	F3161	5-gggagcccagagtccccatgg-3' {SEQ ID
	NO: 156}		
		R3161	5-gggcctccttgggtttgctgg-3' {SEQ ID
	NO: 157}		
15	15	F3541	5-gcctccccagcatcctgccgg-3' {SEQ ID
	NO: 158}		
	•	R3541	5-tcactgagccgaatgaaactgagg-3' {SEQ
	ID NO: 15	9}	
	16	F3531	5-tgtggcctgagttcctttcctgtg-3' {SEQ ID
20	NO: 160}		
	,	R3531	5-ggtcaaagggcagaacgaagaggg-3' {SEQ ID
	NO: 161}		- 33
	-	F3151	5-cccgtccttctcccagccatg-3/ {SEO ID
	17	F3151	5-cccgtccttctcccagccatg-3' {SEQ ID
25	-		
25	17 NO: 162}	F3151 R3151	5-cccgtccttctcccagccatg-3' {SEQ ID 5-ctcccctggttgtccccaagg-3' {SEQ ID
25	17 NO: 162} NO: 163}	R3151	5-ctcccctggttgtccccaagg-3' {SEQ ID
25	17 NO: 162} NO: 163}		
25	17 NO: 162} NO: 163}	R3151 F3141	5-ctcccctggttgtccccaagg-3' {SEQ ID 5-cgacccctctgattgccacttgtg-3' {SEQ ID
	17 NO: 162} NO: 163} 18 NO: 164}	R3151	5-ctcccctggttgtccccaagg-3' {SEQ ID
	17 NO: 162} NO: 163} 18 NO: 164} NO: 165}	R3151 F3141 R3141	5-ctcccctggttgtccccaagg-3' {SEQ ID 5-cgacccctctgattgccacttgtg-3' {SEQ ID 5-ggcatcctgcccttgccaggg-3' {SEQ ID
	17 NO: 162} NO: 163} 18 NO: 164} NO: 165}	R3151 F3141	5-ctcccctggttgtccccaagg-3' {SEQ ID 5-cgacccctctgattgccacttgtg-3' {SEQ ID
	17 NO: 162} NO: 163} 18 NO: 164} NO: 165}	R3151 F3141 R3141 F3522	5-ctcccctggttgtccccaagg-3' {SEQ ID  5-cgacccctctgattgccacttgtg-3' {SEQ ID  5-ggcatcctgcccttgccaggg-3' {SEQ ID  5-tctgtctcccctgctccttg-3' {SEQ ID NO:
	17 NO: 162} NO: 163} 18 NO: 164} NO: 165}	R3151 F3141 R3141	5-ctcccctggttgtccccaagg-3' {SEQ ID 5-cgacccctctgattgccacttgtg-3' {SEQ ID 5-ggcatcctgcccttgccaggg-3' {SEQ ID

- 33 -

	20	F3121	5-gctcctcccgtgaccctctgg-3' {SEQ ID
	NO: 103}		
		R3121	5-gggtcccagccaggagcactg-3' {SEQ ID
_	NO: 104}	T0.14.1	5
5	21	F3111	5-cagcgctcaggcccgtctctc-3' {SEQ ID
	NO: 168}	D2111	E tenatagentatagenenttage 3/ (CEO ID
	NO. 160	R3111	5-tgcataggcatgtgcagctttggg-3' {SEQ ID
	NO: 169}	F3512	5-catgcaccctctgccctgtgg-3' {SEQ ID
10	NO: 170}	F3312	5-catgeaccetetgecetgtgg-5 (5EQ 1D
10	110: 170)	R3512	5-agttgagccaggagaggtggg-3' {SEQ ID
	NO: 171}		
	23	F3101	5-catcaggcgcattccatctgtccg-3' {SEQ ID
	NO: 172}		
15		R3091	5-agcaggagagcagaagaaagg-3' {SEQ ID
	NO: 173}		
	24	F3082	5-gtgtgtcaccatccccaccccg-3' {SEQ ID
	NO: 174}		
		R3082	5-caagagatgggagaaaggccttatg-3' {SEQ
20	ID NO:175	}	
	25	F3073	5-ctgggacatccggatcctgaagg-3' {SEQ ID
	NO: 176}		
		R3073	5-tccaggtagtgggaggcagagg-3' {SEQ ID
	NO: 177}	T2061	5 4
25	26	F3061	5-tcccactacctggagctgccttgg-3' {SEQ
	ID NO: 17	R3051	5-ggctctcccagccctccctg-3' {SEQ ID
	NO: 179}	K3031	3-ggccccccagcccccg-3 (3EQ 1D
	•	F3601	5-cagagcagcagagactctgaccag-3' {SEQ
30	ID NO: 18		· · · · · · · · · · · · · · · · · · ·
		R3601	5-tagaccccacctgcccctgag-3' {SEQ ID
	NO: 181}		
	•	F3501	5-tcctctcattgcttgcctgttcgg-3' {SEQ
	ID NO: 18	2}	·

- 34 -

	R3501	5-ttgagagcttgccggggatgg-3' {SEQ ID				
	NO: 183}					
	29 F3031	5-aagtgccaagcaatgagtgaccgg-3' {SEQ				
	ID NO: 184}					
5	R3021	5-ctcactcccacccacctg-3' {SEQ ID				
	NO: 185}					
	30 F3011	5-cccaccggcctctgagtctgc-3' {SEQ ID				
	NO: 186}	•				
	R3001	5-accctacccaagccaggacaagtg-3' {SEQ				
10	ID NO: 187}					
	31 F2141	5-gaatctgccataaccagcttcgtg-3' {SEQ				
	ID NO: 188}					
	R2141	5-tatcaccccatagaggcctcgaag-3' {SEQ				
	ID NO: 189}					
15	32 F2981	5-cagccactcactctggcacctctg-3' {SEQ				
	ID NO: 190}					
	R2981	5-agcccacagtctctgactctcctg-3' {SEQ				
	ID NO: 191}					
	33 F2131	5-acateteteagggteeetgetgtg-3' {SEQ				
20	ID NO: 192}					
	R2211	5-cctgtgagggacgaggcagg-3' {SEQ ID				
	NO: 193}					
	34 F2202	5-gccctgggtaagggatgctgattc-3' {SEQ				
	ID NO: 194}					
25	R2202	5-cctgcctgggcctcctggatc-3' {SEQ ID				
	NO: 195}	-				
	35 F2111	5-gagggtgatgggggccttagg-3' {SEQ ID				
	NO: 196}	5				
2.0	R2112	5-gcaatcagtttgaagaaggaaagg-3' {SEQ				
30	ID NO: 197}					
	36 F2102 ID NO: 105}	5-cccctctcaccatctcctgatgtg-3' {SEQ				
	R2111	5-agattanaattaaatatanata 2/ (aga				
	ID NO: 106}	5-ggcttcaccttccctctacctcgg-3' {SEQ				
	TO NO. 100)					

- 35 -

	37	F2101	5-cacctttgtctccattctacctgc-3	s' {SEQ
	ID NO: 19	8}		
		R2101	5-ctcccagccccacgcccagg-3'	{SEQ ID
	NO: 199}			_
5	38	F2091	5-ctgagccactctcctcattctgtg-3	3' {SEQ
	ID NO: 20	0}		
		R2091	5-tggaaggggacagtagggagg-3'	{SEQ ID
	NO: 201}			
	39	F2081	5-ggccagtgcgttcttcctcctc-3'	{SEQ ID
10	NO: 202}			
	,	R2071	5-tccctgacctgcccatcatctc-3'	{SEO ID
	NO: 203}		<b>5</b>	. ~
	•	F2061	5-gcccctgtcaggcctggatgg-3'	{SEQ ID
	NO: 204}	12001	3 90000090009900099	(31 913)
a ==	_	D2061	E tenegacian and analysis of	(CEO ID
15		R2061	5-tgacccaggcctccctggagg-3'	{SEQ ID
	NO: 205}			
		F2051	5-ctgaaatggtctctttctttctac-3	3' {SEQ
	ID NO: 20	6}		
		R2051	5-cacaccgactgtcagactgaagag-3	3' {SEQ
20	ID NO: 20	7 }		
	42	F2041	5-ttgtcccctcctaatccccatg-3	3′ {SEQ
	ID NO: 20	8}		
		R2041	5-gggttagggacgtcttcgagg-3'	{SEQ ID
	NO: 209}			
25	43	F2031	5-cagccaaaccatatcaacaatg-3'	{SEQ ID
	NO: 210}		_	-
	,	R2021	5-ctggggaggtgagggctctag-3'	{SEO ID
	NO: 211}			( ~
	44	F2011	5-gaagtgttttgtctcctcctc-3'	(SEO ID
2.0	NO: 212}	12011	3 gaagegeeeegeeeee 3	(52,02)
30	NO: 212 j	D0011	F	(CEO ID
	MO 0101	R2011	5-gcaggcagccagccccatc-3'	SEQ ID
	NO: 213}	D1 602		(050 55
	45	F1021	5-gggtgccctgtgttggctgac-3'	(SEQ ID
	NO: 214}			

- 36 -

```
R1031
                        5-gcaggcagccagccccatc-3' {SEQ ID
   NO: 215}
         46
              F1041
                        5-ctcgtctatgtcttgtgcttgctc-3'
   ID NO: 216}
 5
                        5-caccatggtttggggtcatgtgg-3' {SEQ ID
              R1051
   NO: 217}
        47
                        5-tctcgcttccccagctcctgc-3' {SEQ ID
              F1061
   NO: 218}
              R1061
                        5-tctggagttcgaggactctggg-3' {SEQ ID
10 NO: 219}
        48
              F1071
                        5-agaagggtggggagagaacgg-3' {SEQ ID
   NO: 220}
              R1071
                        5-cagctcagagcctgtggctgg-3' {SEQ ID
   NO: 221}
15
        49
              F1082
                        5-aaggccttcccatcctttggtagg-3' {SEQ
    ID NO: 222}
              R1082
                        5-acaacccagagggagcacggg-3' {SEQ ID
   NO: 223}
         50
              F1092
                        5-gttgacgatgtatatactgtgttgg-3' {SEQ
20 ID NO: 224}
                        5-gctggcaccacagggaatcgg-3' {SEQ ID
              R1091
   NO: 110}
                        5-gcctctctctaactttgcttccttg-3' {SEQ
         51
              F1102
    ID NO: 225}
25
                        5-agccccatgtgcagaatggg-3' {SEQ ID
              R1101
   NO: 112}
         52
              F1112
                        5-ggctacaggctggcagtgatcgag-3' {SEQ
    ID NO: 226}
              R1112
                        5-ttcccccatgcctccactgg-3' {SEQ ID
30 NO: 227}
         53
              F1121
                        5-agccttcgtgcccctaaccaagtg-3' {SEQ
    ID NO: 228}
              R1121
                        5-ctgtgggcattggggctcagg-3' {SEQ ID
    NO: 229}
```

- 37 -

54 F1141 "5-ggatgcccagttgactccggg-3' {SEQ ID NO: 115} 5-cccaccacagtgtcgtcagg-3' {SEO ID R1141 NO: 116} 5-gcccagtgggatcaccatg-3' {SEQ ID 55 F1151 NO: 230} 5-atgctggagggaccccacgg-3' {SEO ID R116 NO: 231}

#### Comparison of Dysferlin With Other Proteins

The 6,243 bp ORF of this candidate MM gene is predicted to encode 2,080 amino acids (Figs. 1C and 2; SEQ ID NO:2). At the amino acid level, this protein is highly homologous to the nematode (Caenorhabditis elegans) protein fer-1 (27% identical, 57% identical or similar: the sequence alignment and comparison was performed using http://vega.igh.cnrs.fr/bin/nph-align\_query.pl.) (Argon & Ward, 1980, Genetics 96:413-33; Achanzar & Ward, 1997, J. Cell Science 110:1073-81). This dystrophy-associated, fer-1-like protein has therefore been designated "dysferlin."

The fer-1 protein was originally identified through molecular genetic analysis of a class of fertilization-defective *C. elegans* mutants in which spermatogenesis is abnormal (Argon & Ward, 1980, *Genetics* 96:413-33). The 25 mutant fer-1 spermatozoa have defective mobility and show imperfect fusion of membranous organelles (Ward et al., 1981, *J. Cell Bio*. 91:26-44). Like fer-1, dysferlin is a large protein with an extensive, highly charged hydrophilic region and a single predicted membrane spanning region at the carboxy terminus (Fig. 3). There is a membrane retention sequence 3' to the membrane spanning stretch, indicating that the protein may be preferentially targeted to either endoplasmic or sarcoplasmic reticulum, probably as a Type II protein

(i.e. with the  $\mathrm{NH_2}$  end and most of the following protein located within the cytoplasm) (Fig. 1C). Several nuclear membrane targeting sequences are predicted within the cytoplasmic domain of the protein

5 (http://psort.nibb.ac.jp/form.html). Immunocytochemical detection of dysferlin suggests that dysferlin is targeted to or anchored within the sarcoplasmic reticulum.

The cytoplasmic component of this protein contains

10 four motifs homologous to C2 domains. C2 domains are
intracellular protein modules composed of 80 - 130 amino
acids (Rizo & Sudhof, 1998, J. Biol. Chem. 273:15897).

Originally identified within a calcium-dependent isoform
of protein kinase C (Nishizuka, 1988, Nature 334:661-65),

- 15 C2 domains are present in numerous proteins. These domains often arise in approximately homologous pairs described as double C2 or DOC2 domains. One DOC2 protein, DOC2α, is brain specific and highly concentrated in synaptic vesicles (Orita et al., 1995, Biochem.
- Biophys. Res. Comm. 206:439-48), while another, DOC2β, is ubiquitously expressed (Sakaguchi et al., 1995, Biochem. Biophys. Res. Comm. 217:1053-61). Many C2 modules can fold to bind calcium, thereby initiating signaling events such as phospholipid binding. At distal nerve
- 25 terminals, for example, the synaptic vesicle protein synaptotagmin has two C2 domains that, upon binding calcium, permit this protein to interact with syntaxin, triggering vesicle fusion with the distal membrane and neurotransmitter release (Sudhof & Rizo, 1996, Neuron

30 17:379-88).

The four dysferlin C2 domains are located at amino acid positions 32-82, 431-475, 1160-1241, and 1582-1660 (Figs. 1C and 3). Indeed, it is almost exclusively through these regions that dysferlin has homology to any 35 proteins other than fer-1. Each of these segments in

- 39 -

dysferlin is considerably smaller than a typical C2
 domain. Moreover, these segments are more widely
 separated in comparison with the paired C2 regions in
 synaptotagmin, DOC2α and β and related C2-positive

5 proteins. For this reason, it is difficult to predict
 whether the four relatively short C2 domains in dysferlin
 function analogously to conventional C2 modules. That
 dysferlin might, by analogy with synaptotagmin, signal
 events such as membrane fusion is suggested by the fact
10 that fer-1 deficient worms show defective membrane
 organelle fusion within spermatozoa (Ward et al., 1981,
 J. Cell Bio. 91:26-44).

The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

#### **EXAMPLES**

Standard methods can be used to synthesize either

#### Example 1: Production of dysferlin protein

wild type or mutant dysferlin, or fragments of either.

20 These methods can also be used to synthesize brainspecific dysferlin polypeptides including full-length or
fragments (e.g., a polypeptide unique to brain-specific

dysferlin). For example, a recombinant expression vector encoding dysferlin (or a fragment thereof: e.g.,

25 dysferlin minus its membrane-spanning region) operably linked to appropriate expression control sequences can be used to express dysferlin in a prokaryotic (e.g., E.coli) or eukaryotic host (e.g., insect cells, yeast cells, or mammalian cells). The protein is then purified by

30 standard techniques. If desired, DNA encoding part or all of the dysferlin sequence can be joined in-frame to DNA encoding a different polypeptide, to produce a chimeric DNA that encodes a hybrid polypeptide. This can be used, for example, to add a tag that will simplify

35 identification or purification of the expressed protein,

- 40 -

or to render the dysferlin (or fragment thereof) more immunogenic.

The preferred means for making short peptide fragments of dysferlin is by chemical synthesis. These fragments, like dysferlin itself, can be used to generate antibodies, or as positive controls for antibody-based assays.

Fusion proteins are useful, e.g., for generating antibodies. Such fusion proteins are generated using 10 known methods. In one example, to construct glutathione S-transferase (GST): dysferlin fusion proteins, the BLAST program (Altschul et al., 1990, J. Molec. Biol. 215:403-410) was used to identify three regions of the dysferlin cDNA that show no homology to any known human proteins 15 (Figure 1). These were subcloned from the dysferlin cDNA as BstYI (881-1333), XmnI (1990-2718) and SalI (5364-5732) fragments ligated respectively into BamHI, SmaI and SalI sites of pGEX-5X-3 (Pharmacia). The three fragments correspond to amino acid sequences at amino acid 20 locations 253-403, 624-865, and 1664-1786 of SEQ ID NO:2, respectively. The resulting GST fusion proteins of BamHI (43 kDa) and SmaI (53.3 kDa) formed isoluble aggregates that were isolated by SDS-PAGE. The fusion protein of SalI (40.2 kDa) was soluble and thus could be purified 25 using a glutathione Sepharose 4B column; the SalI dysferlin fragment (14.2 kDa) was isolated by cleavage from GST using Factor Xa protease. The eluted protein was concentrated and further purified by SDS-PAGE. all three of the fusion peptides, the resulting SDS-PAGE 30 bands were excised and used to immunize rabbits.

### Example 2: Production and characterization of antidysferlin antibodies

Techniques for generating both monoclonal and polyclonal antibodies specific for a particular protein

- 41 -

are well known. The antibodies can be raised against a short peptide epitope of dysferlin, an epitope linked to a known immunogen to enhance immunogenicity, a long fragment of dysferlin, or the intact protein. Antibodies can also be raised against brain-specific dysferlin polypeptides, e.g., against amino acids 1-24 of SEQ ID NO:233. Such antibodies raised against dysferlin or brain-specific dysferlin polypeptides are useful for e.g., localizing such polypeptides in tissue sections or fractionated cell preparations and diagnosing dysferlin-related disorders.

An isolated dysferlin protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind dysferlin using standard techniques 15 for polyclonal and monoclonal antibody preparation. dysferlin immunogen can also be a mutant dysferlin or a fragment of a mutant dysferlin. A full-length dysferlin protein can be used or, alternatively, antigenic peptide fragments of dysferlin can be used as immunogens. 20 antigenic peptide of dysferlin comprises at least 8 (preferably 10, 15, 20, or 30) amino acid residues of the amino acid sequence shown in SEQ ID NO:2 and encompasses an epitope of such that an antibody raised against the peptide forms a specific immune complex with dysferlin. 25 Preferred epitopes encompassed by the antigenic peptide are regions of dysferlin that are located on the surface of the protein, e.g., hydrophilic regions.

A dysferlin immunogen typically is used to prepare antibodies by immunizing a suitable subject (e.g., 30 rabbit, goat, mouse or other mammal) with the immunogen. An appropriate immunogenic preparation can contain, for example, recombinantly expressed dysferlin protein or a chemically synthesized dysferlin polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar

immunostimulatory agent. Immunization of a suitable subject with an immunogenic dysferlin preparation induces a polyclonal anti-dysferlin antibody response.

Polyclonal anti-dysferlin antibodies ("dysferlin 5 antibodies") can be prepared as described above by immunizing a suitable subject with a dysferlin immunogen. The dysferlin antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using 10 immobilized dysferlin. If desired, the antibody molecules directed against dysferlin can be isolated from the mammal (e.g., from the blood) and further purified by well-known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after 15 immunization, e.g., when the dysferlin antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein 20 (1975) Nature 256:495-497, the human B cell hybridoma technique (Kozbor et al. (1983) Immunol. Today 4:72), the EBV-hybridoma technique (Cole et al. (1985), Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for 25 producing hybridomas is well known (see generally Current Protocols in Immunology (1994) Coligan et al. (eds.) John Wiley & Sons, Inc., New York, NY). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with a 30 dysferlin immunogen as described above, and the culture supernatants of the resulting hybridoma cells are

Any of the many well known protocols used for fusing 35 lymphocytes and immortalized cell lines can be applied

antibody that binds dysferlin.

screened to identify a hybridoma producing a monoclonal

- 43 -

for the purpose of generating a monoclonal antibody against dysferlin (see, e.g., Current Protocols in Immunology, supra; Galfre et al. (1977) Nature 266:55052; R.H. Kenneth, in Monoclonal Antibodies: A New Dimension 5 In Biological Analyses, Plenum Publishing Corp., New York, New York (1980); and Lerner (1981) Yale J. Biol. Med., 54:387-402. Moreover, the one in the art will appreciate that there are many variations of such methods which also would be useful. Hybridoma cells producing a 10 monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind dysferlin, e.g., using a standard ELISA assay.

Alternative to preparing monoclonal antibody-15 secreting hybridomas, a monoclonal dysferlin antibody can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (e.g., an antibody phage display library) with dysferlin to thereby isolate immunoqlobulin library members that bind dysferlin. Kits 20 for generating and screening phage display libraries are commercially available (e.g., the Pharmacia Recombinant Phage Antibody System, Catalog No. 27-9400-01; and the Stratagene SurfZAP™ Phage Display Kit, Catalog No. 240612). Additionally, examples of methods and reagents 25 particularly amenable for use in generating and screening antibody display library can be found in, for example, U.S. Patent No. 5,223,409; PCT Publication No. WO 92/18619; PCT Publication No. WO 91/17271; PCT Publication No. WO 92/20791; PCT Publication No. WO 30 92/15679; PCT Publication No. WO 93/01288; PCT Publication No. WO 92/01047; PCT Publication No. WO 92/09690; PCT Publication No. WO 90/02809; Fuchs et al. (1991) Bio/Technology 9:1370-1372; Hay et al. (1992) Hum. Antibod. Hybridomas 3:81-85; Huse et al. (1989) Science

- 44 -

246:1275-1281; Griffiths et al. (1993) *EMBO J*. 12:725-734.

As an example, two polyclonal antisera were raised for each of the fusion peptide antigens described above 5 using New Zealand White rabbits. The rabbits were injected with 0.5 mg of antigen using keyhole limpet hemocyanin (KLH) as the adjuvent. Booster injections of 0.25 mg antigen were administered every three weeks over 12 weeks. Serum was prepared from the rabbits and was purified using affinity column chromatography (HiTrap; Pharmacia) or antigen-blotted polyvinylidene difluoride (PVDF) membrane.

Immunoblotting was used to verify that the affinitypurified antisera recognize the cognate fusion peptides

15 by Western immunoblotting (WIB) and that this reactivity
was immunoadsorbed by pre-incubation of the antisera with
the peptides. Thus, antiserum raised against the
polypeptide encoded by the SalI fragment (encoding amino
acids 1664-1786) identified the fragment both as a

20 cleaved, 14.2 kDa fragment and as a component of the 40.2
kDa GST-SalI fusion peptide. No reactivity was evident
in the fraction containing only the GST fusion partner.
Immunoadsorption entirely abolished this staining.
Analogous results were detected with all six antisera (to

25 the three different target fusion peptides).

#### Preparation of subcellular fractions

Frozen human muscle (0.3 g) was homogenized in five volumes of 0.25 M sucrose containing proteinase inhibitor (Complete, Boehringer). Subcellular fractions of nuclei, 30 mitochondria, microsomes, and cytosol were separated by differential centrifugation. The purity of each fraction was evaluated by immunoblotting of fraction-specific proteins with antibodies to histone H1 (Calbiochem), cytochrome c (Santa Cruz), Na\*-K\* ATPase α1 subunit

- 45 -

(Research Diagnostics) and cytosolic superoxide dismutase (Calbiochem).

#### Dysferlin in subcellular fractions

Immunoblotting was used to analyze dysferlin 5 expression. Twenty  $\mu g$  of each subcellular fraction and 40  $\mu$ q of whole homogenate of muscle were separated by SDS-PAGE (4-15% gradient gel) and transferred to a nitrocellulose membrane. Immunoblotting was performed according to standard methods, using chemiluminescence 10 (ECL, Amersham). Immunoblotting of multi-tissue blots identified prominent dysferlin positively at approximately 230 kDa in heart, placenta, skeletal muscle and kidney. Little or no immuno-positive staining was detected in brain, liver, spleen, ovary, or testis. 15 Lower molecular weight bands (approximately 40 kDa) were also evident. Immunoadsorption with the corresponding fusion peptide abolished both the large and the smaller bands. The 230 kDa band was observed with all of the affinity purified, anti-dysferlin antisera.

Immunoblotting of fractionated human muscle documented distinct 230 kDa bands in the whole muscle homogenate an in microsomal and nuclear fractions. Some immunoreactivity was also evident in the nuclear and mitochondrial fractions. No immunoreactivity was detected in the cytosolic fractions. This pattern was seen with all of the anti-dysferlin antisera, and was eliminated by immunoadsorption. The identity of the assayed fractions was verified by Western blotting using fraction-specific antibodies: histone HI for the nuclear fraction, cytochrome c for the mitochondrial fraction, Na\*-K\* ATPase α1-subunit for the microsomal fraction, and SOD1 for the cytosolic fraction.

#### Example 3: Diagnosis

The discovery of mutations in the dysferlin gene that are associated with the MM and LMGD2B phenotypes means that individuals can be tested for the disease gene before symptoms appear. This will permit genetic testing and counseling of those with a family history of the disease. Additionally, individuals diagnosed with the genetic defect can be closely monitored for the appearance of symptoms, thereby permitting early intervention, including genetic therapy, as appropriate.

10 Individuals with a brain-specific dysferlin-related disorder can be diagnosed using such methods.

Diagnosis can be carried out on any suitable genomic DNA sample from the individual to be tested. Typically, a blood sample from an adult or child, or a sample of placental or umbilical cord cells of a newborn would be used; alternatively, one could utilize a fetal sample obtained by amniocentesis or chorionic villi sampling.

It is expected that standard genetic diagnostic methods can be used. For example, PCR can be utilized to 20 identify the presence of a deletion, addition, or substitution of one or more nucleotides within any one of the exons of dysferlin. Following the PCR reaction, the PCR product can be analyzed by methods such as a heteroduplex detection technique based upon that of White 25 et al. (1992, *Genomics* 12:301-06), or by techniques such as cleavage of RNA-DNA hybrids using RNase A (Myers et al., 1985, Science 230:1242-46), single-stranded conformation polymorphism (SSCP) analysis (Orita et al., 1989, Genomics 10:298-99), di-deoxy-fingerprinting (DDF) 30 (Blaszyk et al., 1995, Biotechniques 18: 256-260) and denaturing gradient gel electrophoresis (DGGE; Myers et al., 1987, Methods Enzymol. 155:501-27). The PCR may be carried out using a primer which adds a G+C rich sequence (termed a "GC-clamp") to one end of the PCR product, thus 35 improving the sensitivity of the subsequent DGGE

- 47 -

procedure (Sheffield et al., 1989, Proc. Natl. Acad. Sci. USA 86:232-36). If the particular mutation present in the patient's family is known to have removed or added a restriction site, or to have significantly increased or decreased the length of a particular restriction fragment, a protocol based upon restriction fragment length polymorphism (RFLP) analysis (perhaps combined with PCR) may be appropriate.

The apparent genetic heterogeneity resulting in the

MM/LGMD2B phenotypes means that the nature of the

particular mutation carried by affected individuals in

the patient's family may have to be ascertained prior to

attempting genetic diagnosis of the patient.

Alternatively, a battery of tests designed to identify

any of several mutations known to result in MM/LGMD2B may

be utilized to screen individuals without a defined

familial genotype. The analysis can be carried out on

any genomic DNA derived from the patient, typically from

a blood sample.

Instead of basing the diagnosis on analysis of the 20 genomic DNA of a patient, one could seek evidence of the mutation in the level or nature of the relevant expression products. Well-known techniques for analyzing expression include mRNA-based methods, such as Northern 25 blots and in situ hybridization (using a nucleic acid probe derived from the relevant cDNA), and quantitative PCR (as described in St-Jacques et al., 1994, Endocrinology 134:2645-57). One could also employ polypeptide based methods, including the use of 30 antibodies specific for the polypeptide of interest. These techniques permit quantitation of the amount of expression of a given gene in the tissue of interest, at least relative to positive and negative controls. One would expect an individual who is heterozygous for a 35 genetic defect affecting the level of expression of

dysferlin to show up to a 50% loss of expression of this gene in such a hybridization or antibody-based assay. antibody specific for the carboxy terminal end would be likely to pick up (by failure to bind to) most or all 5 frameshift and premature termination signal mutations, as well as deletions of the carboxy terminal sequence. of a battery of monoclonal antibodies specific for different epitopes of dysferlin would be useful for rapidly screening cells to detect those expressing mutant 10 forms of dysferlin (i.e., cells which bind to some dysferlin-specific monoclonal antibodies, but not to others), or for quantifying the level of dysferlin on the surface of cells. One could also use a protein truncation assay (Heim et al., 1994, Nature Genetics 15 8:218-19) to screen for any genetic defect which results in the production of a truncated polypeptide instead of the wild type protein.

# <u>Use of immunodetection to identify normal and disease-associated dysferlin</u>

In the following example, immunodetection methods are used to demonstrate a detectable difference in muscles homogenates between normal and disease-associated dysferlin alleles.

Frozen muscle samples (quadriceps) were homogenized in ten volumes of SDS-PAGE sample buffer and boiled for 5 minutes. The final loading volume of SDS-PAGE was adjusted after densitometric measurements (NIH Image) of myosin heavy chain on the Coomassie blue stained gels. Studies were performed on six MM, two LGMD-2B, and three normal muscle samples.

Immunocytochemistry was performed on 8 micron cryostat sections of the muscle that were fixed in 100% cold acetone for 5 minutes and preincubated with PBS containing 1% BSA, 5% heat-inactivated goat serum and 0.2% Triton®X-100. The sections were incubated with

- 49 -

primary antibodies overnight at 4°C and fluoresceinlabeled secondary (TAGO Immunologicals) for 30 minutes at room temperature. The primary antibodies were applied in two double staining combinations: SalI-1 anti-dysferlin and anti-dystrophin antibodies, and SalI-2 anti-dysferlin and anti- $\delta$ -sarcoglycan antibodies. The sections were mounted in SlowFade (Molecular Probes).

The 230 kDA antigen was absent in samples from all five MM patient in immunoblot assays. All five patients 10 had normal patterns of dystrophin expression. Genetic analysis of the dysferlin gene in the patients predicted that at least two of the five MM patients should have no full-length protein. Two of the other three patients had mutations in at least one allele that are predicted to eliminate normal dysferlin expression. In all five patients, absence of dysferlin immuno-staining was documented with at least two other anti-dysferlin antisera.

Immunostaining of dysferlin, dystrophin and  $\delta$ 20 sarcoglycan proteins demonstrated distinct membraneassociated positivity for each protein in normal muscle.
By contrast, in both MM and LGMD-2B muscle the dysferlin protein was absent, while the dystrophin and  $\delta$ sarcoglycan proteins appeared normal.

#### 25 Therapeutic Treatment

A patient with MM/LGMD2B, or an individual genetically susceptible to contracting one or both of these diseases, can be treated by supplying dysferlin therapeutic agents of the present invention. Dysferlin therapeutic agents include a DNA or a subgenomic polynucleotide coding for a functional dysferlin protein. A DNA (e.g., a cDNA) is prepared which encodes the wild type form of the gene operably linked to expression control elements (e.g., promoter and enhancer) that

- 50 -

induce expression in skeletal muscle cells or any other
affected cells. The DNA may be incorporated into a
vector appropriate for transforming the cells, such as a
retrovirus, adenovirus, or adeno-associated virus. One

5 of the many other known types of techniques for
introducing DNA into cells in vivo may be used (e.g.,
liposomes). Particularly useful would be naked DNA
techniques, since naked DNA is known to be readily taken
up by skeletal muscle cells upon injection into muscle.

10 Wildtype dysferlin protein can also be administered to an
individual who either expresses mutant dysferlin protein

Wildtype dysferlin protein can also be administered to an individual who either expresses mutant dysferlin protein or expresses an inadequate amount of dysferlin protein, e.g., a MM/LGMD2B patient.

Administration of the dysferlin therapeutic agents 15 of the invention can include local or systemic administration, including injection, oral administration, particle qun, or catheterized administration, and topical administration. Various methods can be used to administer the therapeutic dysferlin composition directly 20 to a specific site in the body. For example, a specific muscle can be located and the therapeutic dysferlin composition injected several times in several different locations within the body of the muscle. therapeutic dysferlin composition can be directly 25 administered to the surface of the muscle, for example, by topical application of the composition. X-ray imaging can be used to assist in certain of the above delivery methods. Combination therapeutic agents, including a dysferlin protein or polypeptide or a subgenomic 30 dysferlin polynucleotide and other therapeutic agents,

Receptor-mediated targeted delivery of therapeutic compositions containing dysferlin subgenomic polynucleotides to specific tissues can also be used.

35 Receptor-mediated DNA delivery techniques are described

can be administered simultaneously or sequentially.

- 51 -

in, for example, Findeis et al. (1993), Trends in
Biotechnol. 11, 202-05; Chiou et al. (1994), Gene
Therapeutics: Methods and Applications of Direct Gene
Transfer (J.A. Wolff, ed.); Wu & Wu (1988), J. Biol.

5 Chem. 263, 621-24; Wu et al. (1994), J. Biol. Chem. 269,
542-46; Zenke et al. (1990), Proc. Natl. Acad. Sci.
U.S.A. 87, 3655-59; Wu et al. (1991), J. Biol. Chem. 266,
338-42.

Alternatively, a dysferlin therapeutic composition

10 can be introduced into human cells ex vivo, and the cells then implanted into the human. Cells can be removed from a variety of locations including, for example, from a selected muscle. The removed cells can then be contacted with the dysferlin therapeutic composition utilizing any

15 of the above-described techniques, followed by the return of the cells to the human, preferably to or within the vicinity of a muscle. The above-described methods can additionally comprise the steps of depleting fibroblasts or other contaminating non-muscle cells subsequent to

20 removing muscle cells from a human.

Both the dose of the dysferlin composition and the means of administration can be determined based on the specific qualities of the therapeutic composition, the condition, age, and weight of the patient, the progression of the disease, and other relevant factors. If the composition contains dysferlin protein or polypeptide, effective dosages of the composition are in the range of about 1 µg to about 100 mg/kg of patient body weight, e.g., about 50 µg to about 50 mg/kg of patient body weight, e.g., about 500 µg to about 5 mg/kg of patient body weight.

Therapeutic compositions containing dysferlin subgenomic polynucleotides can be administered in a range of about 0.1  $\mu g$  to about 10 mg of DNA/dose for local administration in a gene therapy protocol. Concentration

- 52 -

ranges of about 0.1  $\mu$ g to about 10 mg, e.g., about 1  $\mu$ g to about 1 mg, e.g., about 10  $\mu$ g to about 100  $\mu$ g of DNA can also be used during a gene therapy protocol. Factors such as method of action and efficacy of transformation 5 and expression are considerations that will effect the dosage required for ultimate efficacy of the dysferlin subgenomic polynucleotides. Where greater expression is desired over a larger area of tissue, larger amounts of dysferlin subgenomic polynucleotides or the same amounts 10 readministered in a successive protocol of administrations, or several administrations to different adjacent or close tissue portions of for example, a muscle site, may be required to effect a positive therapeutic outcome. In all cases, routine 15 experimentation in clinical trials will determine specific ranges for optimal therapeutic effect.

#### Animal Model

A line of transgenic animals (e.g., mice, rats, guinea pigs, hamsters, rabbits, or other mammals) can be 20 produced bearing a transgene encoding a defective form of dysferlin. Standard methods of generating such transgenic animals would be used, e.g., as described below.

Alternatively, standard methods of producing null
(i.e., knockout) mice could be used to generate a mouse
which bears one defective and one wild type allele
encoding dysferlin. If desired, two such heterozygous
mice could be crossed to produce offspring which are
homozygous for the mutant allele. The homozygous mutant
offspring would be expected to have a phenotype
comparable to the human MM and/or LGMD2B phenotype, and
so serve as models for the human disease.

For example, in one embodiment, dysferlin mutations are introduced into a dysferlin gene of a cell, e.g., a

fertilized oocyte or an embryonic stem cell. Such cells can then be used to create non-human transgenic animals in which exogenous altered (e.g., mutated) dysferlin sequences have been introduced into their genome or 5 homologously recombinant animals in which endogenous dysferlin nucleic acid sequences have been altered. animals are useful for studying the function and/or activity of dysferlin and for identifying and/or evaluating modulators of dysferlin function. As used 10 herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, 15 dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene 20 product in one or more cell types or tissues of the transgenic animal. As used herein, an "homologously recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous dysferlin gene has been altered by homologous 25 recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to completed development of the animal.

A transgenic animal of the invention can be created 30 by introducing a nucleic acid encoding a dysferlin mutation into the male pronuclei of a fertilized oocyte, e.g., by microinjection or retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. A dysferlin cDNA sequence e.g., that of 35 (SEQ ID NO:1 or SEQ ID NO:3) can be introduced as a

- 54 -

transgene into the genome of a non-human animal. Alternatively, a nonhuman homologue of the human dysferlin gene can be isolated based on hybridization to the human dysferlin sequence (e.g., cDNA) and used as a 5 transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, 10 have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, U.S. Patent No. 4,873,191 and in Hogan, Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). 15 Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the mutant dysferlin transgene in its genome and/or expression of the mutant dysferlin mRNA in tissues or cells of the

animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene.

Moreover, transgenic animals carrying a transgene encoding a mutant dysferlin can further be bred to other transgenic animals carrying other transgenes.

To create an homologously recombinant animal, a vector is prepared which contains at least a portion of a dysferlin gene into which a deletion, addition or substitution has been introduced to thereby alter a dysferlin gene. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous dysferlin gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous dysferlin gene is mutated

- 55 -

or otherwise altered (e.g., contains one of the mutations described in Table 2). In the homologous recombination vector, the altered portion of the dysferlin sequence is flanked at its 5' and 3' ends by additional nucleic acid 5 of the dysferlin gene to allow for homologous recombination to occur between the exogenous dysferlin nucleic acid sequence carried by the vector and an endogenous dysferlin gene in an embryonic stem cell. additional flanking dysferlin nucleic acid is of 10 sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see, e.g., Thomas and Capecchi (1987) Cell 51:503 for a description of homologous recombination 15 vectors). The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced dysferlin sequence has homologously recombined with the endogenous dysferlin gene are selected (see, e.g., Li et al. (1992) Cell 69:915). 20 selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras (see, e.g., Bradley in Teratocarcinomas and Embryonic Stem Cells: A Practical Approach, Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be 25 implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline 30 transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley (1991) Current Opinion in Bio/Technology 2:823-829 and in PCT Publication Nos. WO 90/11354, WO 91/01140, WO

BNSDOCID: <WO\_\_\_\_\_0011157A1\_IA>

35 92/0968, and WO 93/04169.

- 56 -

#### Other Embodiments

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

BNSDOCID: <WO\_\_\_\_\_0011157A1\_IA>

- 57 -

What is claimed is:

- 1. An isolated DNA comprising a nucleotide sequence which hybridizes under stringent hybridization conditions to SEQ ID NO:3, or a complement thereof.
- 5 2. The isolated DNA of claim 1, wherein the nucleotide sequence is SEQ ID NO:117.
  - 3. An isolated DNA comprising a nucleotide sequence selected from the group consisting of SEQ ID NOs:4-12.
- 4. The isolated DNA of claim 3, comprising the sequence of SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, or SEQ ID NO:21.
  - 5. An isolated DNA comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:22-30.
- 15 6. A single stranded oligonucleotide of 14-50 nucleotides in length having a nucleotide sequence identical to a portion of SEQ ID NO:3, or a complement thereof.
  - 7. A pair of PCR primers consisting of:
- 20 (a) a first single stranded oligonucleotide consisting of 14-50 contiguous nucleotides that are identical to a portion of SEQ ID NO:117; and
- (b) a second single stranded oligonucleotide consisting of 14-50 contiguous nucleotides that are identical to a portion of SEQ ID NO:117, wherein the sequence of at least one of the oligonucleotides is identical to a portion of a strand of SEQ ID NO:3, and the first oligonucleotide is not complementary to the second oligonucleotide.

- 8. A pair of single-stranded oligonucleotides, wherein both oligonucleotides are selected from the group consisting of SEQ ID NOS:130-231, SEQ ID NO:110, and SEQ ID NO:112 and the oligonucleotides are different from 5 each other.
  - 9. An isolated DNA comprising a nucleotide sequence that encodes a polypeptide that shares at least 70% sequence identity with SEQ ID NO:2, or a complement of the nucleotide sequence.
- 10 10. The isolated DNA of claim 9, wherein the polypeptide comprises the sequence of SEQ ID NO:2.
- 11. An isolated DNA comprising a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid having a sequence selected from the group consisting of SEQ ID NOs:31-79 and 90-100.
- 12. A single stranded oligonucleotide of 14-50 nucleotides in length comprising a nucleotide sequence which is identical to a portion of a nucleic acid selected from the group consisting of SEQ ID NOs:31-79 and 90-100, or a complement of the nucleotide sequence.
  - 13. The oligonucleotide of claim 12, wherein the portion includes an intronic sequence.
    - 14. A pair of PCR primers consisting of:
- (a) a first single-stranded oligonucleotide
  25 consisting of 14-50 contiguous nucleotides that are identical to a portion of a sense strand of a nucleic acid selected from the group consisting of SEQ ID NOs:31-85; and

PCT/US99/19395

- (b) a second single stranded oligonucleotide consisting of 14-50 contiguous nucleotides that are identical to a portion of the antisense strand of a nucleic acid selected from the group consisting of SEQ ID NOs:31-85, wherein the sequence of at least one of the oligonucleotides comprises a sequence identical to a portion of a nucleic acid selected from SEQ ID NOs: 31-79 and 90-100, and wherein the first oligonucleotide is not complementary to the second oligonucleotide.
- 15. A pair of single-stranded oligonucleotides selected from the group consisting of SEQ ID NOs:101-116, SEQ ID NOs:184-185, SEQ ID NOs:188-191, SEQ ID NOs:210-213, and SEQ ID NOs:216-217.
- 16. A vector comprising the isolated DNA of claim15 1.
  - 17. A substantially pure polypeptide comprising an amino acid sequence sharing at least 70% sequence identity with SEQ ID NO:2.
- 18. The substantially pure polypeptide of claim 17, 20 wherein the polypeptide comprises an amino acid sequence identical to that of a naturally occurring polypeptide.
  - 19. The substantially pure polypeptide of claim 18, wherein the amino acid sequence comprises the sequence of SEO ID NO:2.
- 20. A substantially pure polypeptide comprising an amino acid sequence identical to the amino acid sequence of amino acid residues 1-500, 501-1000, 1001-1500, or 1501-2080 of SEQ ID NO:2.

- 21. A substantially pure polypeptide comprising the amino acid sequence of SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88 or SEQ ID NO:89.
- 22. A substantially pure polypeptide selected from 5 the group consisting of amino acids 253-403 of SEQ ID NO:2, amino acids 624-865 of SEQ ID NO:2, and amino acids 1664-1786 of SEQ ID NO:2.
  - 23. A fusion protein comprising a polypeptide of claim 22.
- 10 24. An antibody that specifically binds to the polypeptide of claim 22.
  - 25. An antibody that binds specifically to the polypeptide of claim 17.
    - 26. A cell comprising the isolated DNA of claim 1.
- 15 27. A non-human mammal, the genomic DNA of which bears a transgene, wherein the transgene comprises the isolated DNA of claim 1.
- 28. A transgenic non-human mammal having a transgene disrupting or interfering with the expression 20 of a dysferlin gene.
  - 29. A method of decreasing the symptoms of muscular dystrophy in a mammal, the method comprising introducing into a cell of said mammal the isolated DNA of claim 1.
- 30. A method of decreasing the symptoms of muscular 25 dystrophy in a mammal, the method comprising introducing

- 61 -

into a cell of said mammal the vector of claim 16, the vector being an expression vector.

- 31. A method of decreasing the symptoms of muscular dystrophy in a mammal, the method comprising introducing 5 into a cell of said mammal the protein of claim 17.
  - 32. A method for identifying a patient, a fetus, or a pre-embryo at risk for having a dysferlin-related disorder, the method comprising:
- (a) obtaining a sample of genomic DNA from the10 patient, fetus, or pre-embryo; and
  - (b) determining whether the sample contains a mutation in a dysferlin gene, wherein a patient, a fetus, or a pre-embryo having a mutation in a dysferlin gene is at risk for having a dysferlin-related disorder.
- 15 33. The method of claim 32, comprising:
  - (a) treating the sample of genomic DNA with a restriction enzyme specific for a particular restriction enzyme site; and
- (b) detecting the presence or absence of the 20 particular restriction enzyme site in the sample of genomic DNA as an indication of the presence or absence of a particular mutation in the genomic DNA.
- 34. The method of claim 33, wherein the restriction enzyme is selected from the group consisting of Pst I,
  25 Fnu4H I, BamH I, BstY I, Ava II, HinP I, Fsp I, Mbo II,
  ScrF I, BstN I, Mae I, Bfa I, Dde I, Bpm I, Ban II, Ava
  II, and Sau96 I.
  - 35. The method of claim 32, comprising subjecting the sample to polymerase chain reaction (PCR).

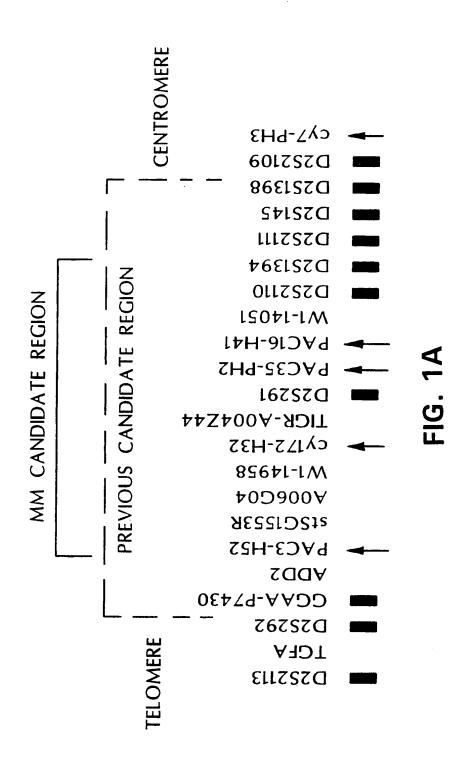
- 36. The method of claim 32, comprising:
- (a) contacting a single stranded oligonucleotide with the sample of genomic DNA; and
- (c) detecting hybridization or lack thereof between 5 the single stranded oligonucleotide and the genomic DNA, as an indication of the presence or absence of a mutation in the genomic DNA.
- 37. A method for identifying a patient, a fetus, or a pre-embryo at risk for having a dysferlin-related10 disorder, said method comprising:
  - (a) providing a sample comprising dysferlin mRNA from the patient, fetus, or pre-embryo; and
- (b) determining whether the dysferlin mRNA contains a mutation, wherein a patient, a fetus, or a pre-embryo15 having a dysferlin mRNA containing a mutation is at risk for having a dysferlin-related disorder.
  - 38. The method of claim 37, wherein the presence or absence of the mutation is detected by Northern blot.
- 39. The method of claim 37, wherein the method 20 includes the step of subjecting the sample to polymerase chain reaction (PCR).
  - 40. A method for detecting the absence of a mutation in a dysferlin protein of a patient, a fetus, or a pre-embryo, the method comprising:
- 25 (a) providing a sample comprising a dysferlin protein of the patient, fetus, or pre-embryo;
  - (b) contacting the sample with the antibody of claim 22; and
- (c) detecting binding of the antibody to dysferlin 30 protein in the sample, if any, wherein binding indicates a normal dysferlin protein.

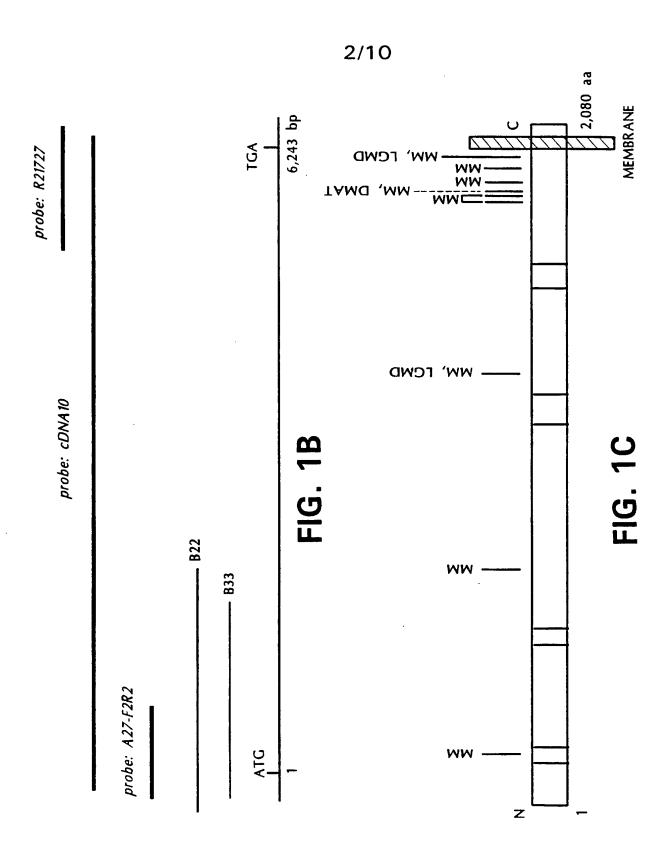
- 63 -

- 41. An isolated DNA comprising a nucleotide sequence that is identical to the sequence of amino acid residues 3501-3520 of SEQ ID NO:1, 3737-3756 of SEQ ID NO:1, 3842-3861 of SEQ ID NO:1, 5114-5139 of SEQ ID NO:1, or 5239-5255 of SEQ ID NO:1.
  - 42. An isolated DNA comprising a nucleotide sequence selected from the group consisting of 3501-3520 of SEQ ID NO:1, wherein nucleotide G at 3510 is A;
- 3737-3756 of SEQ ID NO:1, wherein nucleotide G at 3746 is deleted;
  - 3842-3861 of SEQ ID NO:1, wherein nucleotide C at 3851 is T;
  - 5114-5139 of SEQ ID NO:1, wherein nucleotide C at
- 15 5122 and nucleotide A at 5123 are deleted; 5239-5255 of SEQ ID NO:1, wherein nucleotide G at
  - 5245 is deleted and nucleotide G at 5249 is C; and 5239-5255 of SEQ ID NO:1, wherein nucleotide G at 5245 is C and nucleotide G at 5249 is deleted.
- 20 43. An isolated nucleic acid comprising a nucleotide sequence which hybridizes under stringent hybridization conditions to nucleic acids 3284-3720 of SEQ ID NO:232, or the complement of said nucleotide sequence.
- 25 44. An isolated nucleic acid comprising a nucleotide sequence identical to the sequence of nucleotides 3284-3720 of SEQ ID NO:232, or a complement of said nucleotide sequence.
- 45. The isolated nucleic acid of claim 44, wherein 30 the nucleotide sequence comprises the sequence of SEQ ID NO:232 or the complement of SEQ ID NO:232.

- 46. An isolated polypeptide comprising:
- a) at least 15 contiguous amino acids of the polypeptide comprising amino acids 1-24 of SEQ ID NO:233,
- b) a naturally occurring allelic variant of a
   5 polypeptide comprising amino acids 1-24 of SEQ ID NO:233,
   or
  - c) an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes under stringent conditions to nucleotides 3284-3720 of SEQ ID NO:232.
- 10 47. The polypeptide of claim 46, wherein the polypeptide comprises SEQ ID NO:233.
  - 48. A vector comprising the nucleic acid of claim 44.
    - 49. A cell comprising the vector of claim 48.
- 15 50. A method of making a polypeptide, the method comprising culturing the cell of claim 49.
  - 51. An antibody which specifically binds to a polypeptide of claim 46.
- 52. The antibody of claim 51, wherein the antibody 20 binds to a polypeptide selected from the group comprising amino acids 253-403 of SEQ ID NO:233, amino acids 624-865 of SEQ ID NO:233, and amino acids 1664-1786 of SEQ ID NO:233.
- 53. The antibody of claim 51, wherein the antibody 25 is a monclonal antibody.
  - 54. The antibody of claim 51, wherein the antibody is a polyclonal antibody.

1/10





**SUBSTITUTE SHEET (RULE 26)** 

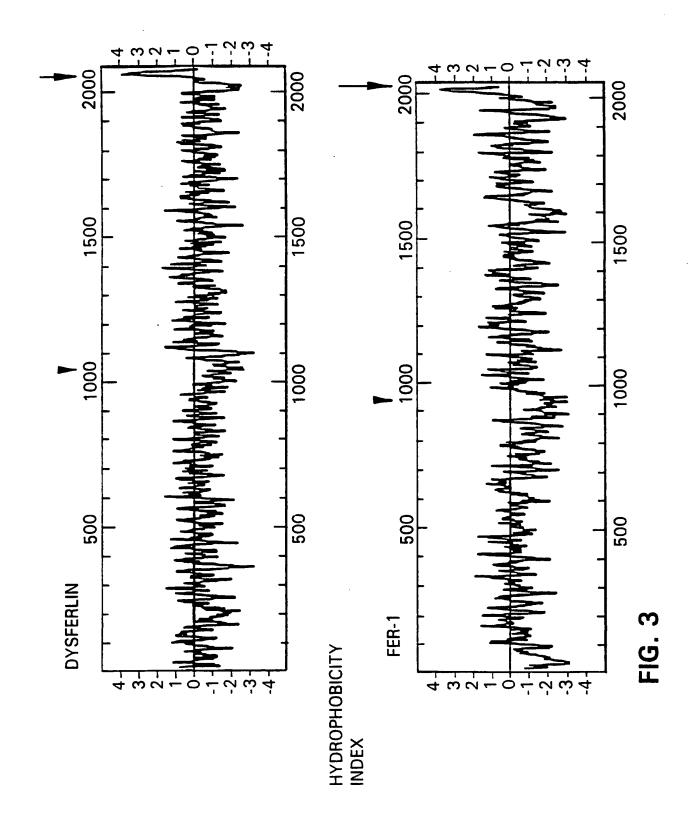
#### 3/10

```
1 MERVEILYAE NUHTPOTDIS DAYCSAVFAG VKKRTKVIKN SUNDUWNEGF
   51 EVOLKGIPLD CGSELHVVVK CHETMGPARF LGEAKVPLRE VLATPSLSAS
  101 FNAPLLDTKK QPTGASLVLQ VSYTPLPGAV PLFPPPTPLE PSPTLPDLDV
  151 VADTGGEEDT EDQGLTGDEA EFFLEQSGGP GAPTTPEXLP SRPPPHYPGI
  201 KRKRSAPTSR KLLSDKPQDF CIRVQVIEGR QLPGVNIKPV VKVTAAGQTK
  251 RTRIHKGNSP LFNETLFFNL FDSPGELFDE PIFITVJDSR SLRTDALLGE
  301 FRMDVGTIYR EPRHAYLRKW LLLSDPDEFS AGARGYLKTS LCVLGPGDEA
  351 PLERKDPSED KEDIESNLLR PTGVALRGAH FCLKVFRAED LPQMDDAVMD
  401 NVKQIFGFES NKKNLVDPFV EVSFAGKMLC SKILEKTANE OWNONITLEA
  451 MEPSMCEWAR IRLIDWDRLT HNDIVATTYL SMSKISAPGG ELEEEPAGAV
  501 KPSKASDLDD YLGFLPTFGP CYINLYGSPR EFTGFPDPYT ELNTGKGEGV
  551 AYRGRLLISL ETKLVEHSEQ KVEDLPADDI LRVEKYLRRR KYSLFAAFYS
  601 ATMLQDVDDA IQFEVSIGNY GNKFDMTCLP LASTTQYSRA VFDGCHYYYL
 651 PWGNVKPVVV LSSYWEDISH RIETQNQLLG IADRLEAGLE QVHLALKAQC
701 STEDVDSLVA QLTDELIAGC SQPLGDIHET PSATHLDQYL YQLRTHHLSQ
 751 ITEAALALKL GHSELPAALE QAEDWLLRLR ALAEEPQNSL PDIVIWMLQG
 801 DKRVAYQRVP AHQVLFSRRG ANYCGKNCGK LQTIFLKYPM EKVPGARMPV
 851 QIRVKLWFGL SVDEKEFNQF AEGKLSVFAE TYENETKLAL VGNWGTTGLT
 901 YPKFSDVTGK IKLPKDSFRP SAGWTWAGDW FVCPEKTLLH DMDAGHLSFV
 951 EEVFENQTRL PGGQWIYMSD NYTDVNGEKV LPKDDIECPL GWKWEDEEWS
1001 TDLNRAVDEQ GWEYSITIPP ERKPKHWVPA EKMYYTERRR RWVRLRRRDL
1051 SQMEALKRER QAFAEGEGWE YASLFGWKFH LEYRKTDAFR RRRWRRRMEP
1101 LEKTGPAAVF ALEGALGGVM DDKSEDSMSV STLSFGVNRP TISCIFDYGN
1151 RYHLRCYMYQ ARDLAAMDKD SFSDRYAIVS FLHOSOKTYV VYOTTLNRTWD
1201 OTLIFYEIEI FGEPATVAEO PPSIVVELYD HDTYGADEFM GRCICOPSLE
1251 RMPRLAWFPL TRGSOPSGEL LASFELIORE KPAIHHIPGF EVQETSRILD
1301 ESEDTDLPYP PPQREANIYM VPQNIKPALQ RTAIEILAWG LRNMKSYQLA
1351 NISSPSLVVE CGGQTVQSCV IRNLRKNPNF DICTLFMEVM LPREELYCPP
1401 ITVKVIDNRQ FGRRPVVGQC TIRSLESFLC DPYSAESPSP QGGPDDVSLL
1451 SPGEDVLIDI DDKEPLIPIQ EEEFIDWWSK FFASIGEREK CGSYLEKDFD
1501 TLKVYDTQLE NVEAFEGLSD FCNTFKLYRG KTQEETEDPS VIGEFKGLFK
1551 IYPLPEDPAI PMPPRQFHQL AAQGPQECLV RIVIVPAFGL OPKDPNGKCD
1501 PYTKISIGKK SYSDODNYIP CTLEDVEGKM FELTCTLELE KOLKITLYDY
1551 DLLSKDEKIG ETVVDLENRL LSKFGARCGL PQTYCVSGPN QWRDQLRPSQ
1701 LLHLFCQQHR VKAPVYRTDR VMFQDKEYSI EEIEAGRIPN PHLGPVEERL
1751 ALHVLQQQGL VPEHVESRPL YSPLQPDIEQ GKLQMWVDLF PKALGRPGPP
1801 FNITPRRARR FFLRCIIWNT RDVILDDLSL TGEKMSDIYV KGWMIGFEEH
1851 KQKTDVHYRS LGGEGNFNWR FIFPFDYLPA EQVCTIAKKD AFWRLDKTES
1901 KIPARVVFQI WDNDKFSFDD FLGSLQLDLN RMPKPAKTAK KCSLDQLDDA
1951 FHPEWFVSLF EQKTVKGWWP CVAEEGEKKI LAGKLEMTLE IVAESEHEER
2001 PAGQGRDEPN MNPKLEDPRR PDTSFLWFTS PYKTMKFILW RRFRWAIILF
2051 IILFILLEL AIFIYAFPNY AAMKLUKEES
                                                (SEQ ID NO:2)
```

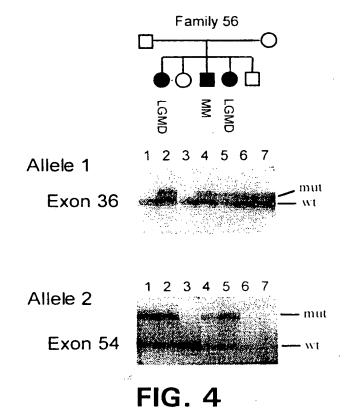
# FIG. 2

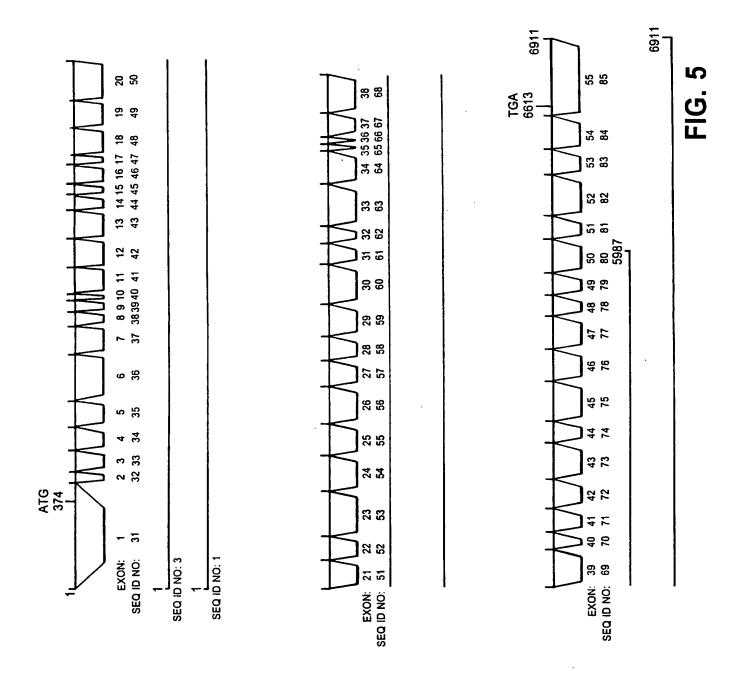
#### SUBSTITUTE SHEET (RULE 26)

4/10



## SUBSTITUTE SHEET (RULE 26)





4		4	Ą	i	æ			5			ည္က			30			30			
	ш	4	ia E	! <u>'</u>	¥	æ		Ö	J		ğ	Ω		gağ	z		c agc	S	<b>6A</b>	
A TITE	æ	CTA ANA	ب م	ິ :	6	U		o o	۲		ο Ο	œ		g aag	×		g C	Д	0	•
	α.	£	ב מ	α,	A.A.	×		ŭ	ഗ		ğ	А О		grg (			8	O.	9	•
		111			GAG AAA GGA	ш		gto	>		S	ø		grg			Cač	o	$\cong$	•
	ဟ	ព្រះព	į		Ë	د.		ដ	ហ		tac	.×		grg	>		gag	(LI)	Ц	•
		AGT S		ы	A	S		atg	Σ		at	Σ		acg			gat	K		
CCG P TTG	<b>ر</b> .	e Gy	440	C	TTG	<u>.</u>		tcc	ល		tac	7.		aag	×		gtt	>		
8 . 8		E.	Ę	}	ð			at	_		S G			ag	_		Ca B			
1 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3	а 18	4 X	111	; ! :2	141	(2)	171	gaa	ய	201	obo	œ	231	agc	۲n	251	Soci			
61/2 CCA ( P 151/	C	424 X	331/		7124	د.	511/	agt	ΕΩ	501/	cra	_ _	591/	cag		731/	gog (	۵.		
61/21 GCT CCA CCA P A P P S 151/51 GAT CTG CCC P	0	8	3	~	233			aag	~		ar (	_		Jac (	~	•	gag (	61		
TAT OF	"		2		200	~		Jac i	_		ac (	_		tg (			) 265	 /B		
8 2 2	·	46,4	E	,,	A E	,		Jat o	_		) jgc	~		ST:	·-		נננ			
TAC AGG CAT O	٠, د	٠٤ ١٤	Ē		ဥ	. <del>.</del> .		atg c			aac o	-		200			atc t			
GAT 1 D )	a L	TAT :	į.		4.5	حت		tg ë	_		966	-		gta	·		gag			
	า		TA	_	TGA GAA	ш		ğ			at			atco			atc			
AGC 1 S W	ы П	AT .	ç	<del>ئە</del>	13			g Sg			act	_		gcc a			gag a	-		
AGT A S S CTC G	បោ	42	12 12	3	SC A	Σ		AG g	G		to g	Ω		at g	∢		tac g			
ST SE	 	S E N	년 년 년	ند	લ કુટ ઇ	>	21	3	ш	91	tat	Ĺŧ.	77	S Č	>-	21	S	>-		
31/11 CTC CCG L P 121/41 GCT GGT	G 11/7	£ 3	1/10	<b>-</b>	1 5 2 7	Ω	31/1	ড গ	G	11/1	jc a	H	1/2	ñ	Ωı	17/5	ü	4		
ស្គី ជ	۲ آگ	ر الا ≪	G G	اد	∺ 2 •	z	37	F.	Δ.	ķ	Ö.	Ü	ĕ	r g	Ω	2,	n at	H		
	ø							A G	Ω		r C	S		ב בכב			gctc	u		
COTC L L GGT		200 200 200 200 200 200 200 200 200 200		z				P AGA				H		: tt			acg (			
555 c GGC	>	υ 20 4	3	X	Ö	ը		5	ပ		acc	H		ដ	S		ည်	o		
s s CAT	<b></b>	ig.	AAG	×	TGA	•		AGA	œ		ü	Δ,		gac	Ω		gac	Ω		
F. T. D.	U	ű E	5	ບ	ဗ္ဗ	G		g	G		age.	α.		aaç	×		rgg	3		
CGA R TTT	ĹĿı	AGT S	TAA		AGA	œ		g	ပ		aac	z		gac	Ω		acc	⊬		
AAG GG	ပ	2 N	AAT	z	G			ð	Ω,		gtg	>		atg	Σ		S	ዉ		
TTC F ACG	۴	ပ္ပ ဗ	AAT	z	TGA	•		AAC	z		ggt	ပ		gcg	æ		aac	z		
7GG W 11 GAG	ы 6	ű ű	15 A	Ę-	GAT	۵	151	919	>	131	tto 1	(L,	211	gat	æ	241	Ctt	L		
1/1 TCC TGG S W 91/31 ATA GAG	I 181/	3. 1.	271./ GGT	ی	361/ GGA	ၒ	451/	EE:	li.	541/	agc	S	631/	crg	<b>.</b>	721/	300	۲		

# **SUBSTITUTE SHEET (RULE 26)**

																						_							
	CCa			ည္ထ	Æ		cag	o		aac	z		aac	z		gcc	>		rac	>-		aag	×		CC CC	Ŋ		aac	<sub>2</sub> [
	atg			5 5 5	۵.		ບ	Ω,		ctg cgg 6	oc.		agg	α;		מנכ מככ	H		CCC	Ω,		gac	Ω		ggc	ၒ		tgt o	ر
	665	<b>~</b>		aag	×		CCa	۵,		ctg			atc	н		arc	н		gac	Ω		gat	۵		tga	U		ו ננ	ı. (
	gaa	ω		gag	ω		E C C C C	D.		ggc	G		gtc	>		SCC	م		tgt	_ ပ		att	H		aag	×		gac	ے ا
	ctg	<b>ر.</b>		aga	~		tac	>+		99			g			CCC	۵,		ctg	د.		gac	Ω		gaa	ш		rot o	درد
	agt	S		cag	o		S	۰۵		gca	æ		t CC	S		tgc	ပ		tto	ĹĿ,		atc	₩		agg	n:		g gcc ttt gag ggc ctg t	_
	ccg	۵.		atc	H		ctg	L.		ctg			cag	0		tac	<b>&gt;</b>		agc	တ		CtC	_		gag	ш		ggc	<del>ن</del>
	caa	o		CEC	J		gac	Ω		atc	_		gtg	>		CtC	.ı		gag	(c)		gtg	>		999	ပ		gag	ĹĹ
291	tgt	U	321	gag	ш	/351	aca	۲	/381	gag	ធា	/411	acg	E	/441	gag	ш	/471	ctg	<u>۔</u>	/501	gac	۵	/531	ata	ы	/561	ttt	ie.
871/29	atc	H	961/321	tt	LL	1051	gac	Ω	1141	atc	⊷	1231	cag	o	1321	gag	ы	1411	too	S	1501	gaa	ш	1591	tcc	S	1681	gcc	∢
	tgc	U		ü	S		gaç	ш		gc	K		ğ			äğ	<u>~</u>		ğ	~		ğ			ğ	4.		ga	ы
	ენე	α.		ეე <u>ნ</u>	æ		tot t	S		acc	⊢		999	ပ		S	Δ,		atc	н		CCt	a,		これ	ĹL,		gtg	>
	ggt	ပ		ctg	L		gag	ш		agt	œ		tgt	ы О		ctg	ار س		acc	T I		agt	S		asa tto ttt	Ĺ		aat	z
	atg	Σ		Ctg	J		gat	Ω		cag	o		gag	ш		atg	Σ		rgt	ن		CtC	ᆈ		aaa	×		gag	(e)
	tt	л Ж		gag	ш		ctg	Ļ		CTC			gta	>		gtg	>		cag	0		cta	LLSP		agc	ຜ		ctg	
	gag	ய		999	<del>ن</del>		atc	н		g			r Q			aa			б			Б	S		tgg	3		cag	
	gac	Ω		tag	S		agg	<b>c</b> .		င္ပင္ခ	۵,		Ctc			atg	Σ		gtg	>		gtg	>		tgg	3		aca	₽
	gca	Æ		Scg	۵.		tca	s		aag	<u>يد</u>		agc	ຮ		tto	Ŀ.		gtg	>		gat	Ω		gat	۵		gac	Ω
281	<b>9</b> 95	ပ	311	cag	ø	1/341	aca	⊣	1/3/1	atc	щ	/401	ပ္ပ	ρ,	1/431	ctc	د	1/46]	CCC	بھ	1/49]	gac	۵	1/52]	atc	н	1/551	tat	<b>&gt;</b> -
841/	tat	>•	931/		S	1021	gag	ш	111	aac	z	1201	ដូ	ທ	1291/431	acc	£-	1381	cgg	œ	147]	CCa	, a a a	156	tte ate gut t	(4.	165	gtc tat gac a	>
	act	E		266	ပ		cag	o		cag	0	,	r C	S		tgc	່ບ		gg	<b>~</b>		ggc			_			-	
	gac	Ω		agg	n.		gtg	>		CCT	۵۰		atc	H		atc	н		ggc	G		ggt	ပ		gaa	ω		ctg	د
	cat	=		acd	•		gag	ы		gtt	>		aac	2	;	gac	۵		ttt	(L,		cag	o		gag	ш		acc	₽
	gac	. Δ		מנק	, 		נננ	(μ,		atg	Σ		gua	, A	:	tt	Ĺ		cag	0	,	CCa	۵.		cag	0	,	gac	Ω
	tac	>4		S	Ω,		qqt	ဗ		tac	>	ı	מ	<u> </u>	,	aac	z		Spo	· œ.		to	S		atc	₩		ttt	L.
	cta	, ب		ננט	L.		act	Ω,		atc	<b>—</b>	ı	ממ	) } }	¥	ບບບ	ىم		aac	Z		cca	م		S	۵.		gat	Ω
	gag	hш	ı	בממ	9 3		att	-	ı	aac	z	:	1	} >		aac	2		gat	۵ د		agt	, S		ato	H		aag	×
	מנמ	, , ,		g			cat	<b>=</b>		S	) A	:	1	ີ ເ	,	aad	)   	<b>.</b>	atc		١.	gag	ы		ctc		١	gag	ш
271	מנט	) ) ) )	301	בנם	n ,	331	Cac		7361	gag	i L	7391		n 1	7421	200	ה ה ה	/451	atc	, >	/481	g	A	/511	S	۵	/541	ctg	د
811/	بر 1	) 	901/	000	ה ה ה	991/	arc	<b>.</b>	1081	add	n	1171		5 2	1261	ָ ער ה ער ה	) 	1351	aad	<b>×</b>	1441	i c	S	1531	939	i i	1621	tac	YLEKDFDTLK

						_		_	_	*1	m	t n
	U U		rac Y	agt S	cta L	aag K	Cac H	ara I	9ag E	A gcc	gtg V	n aag K
	tat	<b>&gt;</b> -	atc I	gtg V	gac D	S	C C C	949	P	aag K	ı gat D	сза Q
	att	ы	cgt R	tca S	aag K	ctg L	ָר ני	gaa E	gtc V	g d	aga R	aag K
	aaa	×	gtc V	aaa ×	gag n	ctg L	cag Q	att	ctg L	cta ttt L F	aat acc N T	gaa cac E H
	בנט	Ĺ.	ttg L	aag K	ctg L	agg R	S	tat too a Y S	990 G	cta	aar N	gaa
	ctc	_	tgc C	999 G	CC T	aac N	<b>9</b>	tar Y	0.39 Q	gac D	tgg W	gaa
	agc	<del>ن</del>	gag E	ata I	acc tgc act ctg c T C T L P 2131/711	gag E	cgc R	a gaa t. E Y	0 0	gtc V	atc I	F F
	aag	F K 1/621	cag Q	t c c	act.	ctg L	ctc L	aaa *	cag Q	tgg W	att I '	atg att ggc t M I G F
/591	ttt	F /621	CCC P	atc I /681	tgc C 7111	gac D /741	cag 0 /771	gat D /801	ctt L /831	atg M /861	rgt C /891	att
1771	gaa ttt a	E 1861	gga ccc c G P Q	aag K	acc T 2131	gtc V 2221	gac D 2311	c39 0 2401	9tg V 2491	cag Q 2581	cgt R 2671	atg M
	<b>3gt</b>	(2	cag O	atc I	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	200	8	<u>.</u>	i m	i ii	בַ נַנ	τ, Ω
	מנה י	_	U Z	rac	gag	gag acg	tgg %	gta atg t V M F	gct ctg A	aag K	rtt ttc F F	ggt G
	tg a	н >	dcc o	P	F	D ag	cag	gta.	gct.	999	יין נד	aaa K
	ict ic		ctg :	gat	arg 4	9gr	aac N	cgt.	otg L	cag O	agg R	979 V
	E C	ss ss	cag Q	tgt c	aag K	מי	600 600	gac	7 gt	ge c	aga	א ני ני
	Jat o		ac	ag.	gga (	aag acc K I (	ga	ខ្ល	ge	atc o	gcc	art tat I Y
	Jaa ç	_ О Ш	cca aga cag ttc c P R Q F H	gga d G	gag ccc gta ttt gr E P V F G 2101/701	Jaa 1	: tac tgt gtc tct g Y C V S G 2281/761	9 660	9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9	ĕ	rg S	ပ္ထ
	Ca C	<u>.</u>	ag t	lat g	Jta (	yac ç	Jtc 1	3000	gtg .	S C C	3 3	280
581	ag a	E T 1/611	98 0	gac ccc aat g D P N G	10 C	ag ç ( 1 (731	:gt ç	159 t γ γ791	ggc cca gtg G P V 2461/821	ctg cag cca g L Q P I 2551/851	: acc cca cgg ag T P R R 2641/881	atg agc g
1741/581	ag g	E E 1831/	Ca a	ac c	ag c E E	cc a	ac t	set 9	3 E	tg 0	cc c	adg a
	_	ω	200 G	aag g K C	ctg 9 L E	ctc t	acc t	gca c	ctg g L C	CCC C	atc a I 1	gag a
	cg C	<del>ا</del>	atg c M P	000 g	acg o	ដូច	ag a	5 6e	) ) (1)	ο B.	ac a 	999 9 G
	ag a	<del>π</del>	o D D D D	cag c	tgc a	on on	g _	ה ה	cca cac P H T	ရ က လ	ttc aac F N	acg g T
			atc c I P	p D O	ccc t P C	e D	ဂ္ဂ . ဂ <del>။</del>	ga g V	ភ្ជ ភ ដ	ה אר	CCC P	
	19 g	O	ğ μ	κ Ω Ή	ប៉ូ <u>o</u> .	מ ≺ הַ	ga C	מר א	ró Z roj			
	ပ္	æ	ĕ. ĕ.	์ ผู้ ถ	tac atc ( Y I	at D	ال 0.0	ο π E	Ω 64	cgg c R P	a D D	ctg a
	ig Ci	<b>&gt;</b> +	ក្ត	e E	گ ۲۰ ټې	אַ יע	n C	g O	99 a	rg rg	ม ม	ភ្ជ
	ם ה	<u> </u>	la gé D	ag A	1981/551 gac cag gat aac t D Q D N \ 2071/691	Ή Ω ¬	ή Ω κ	ပို့ ပို့	بة م	lg to S	D. D.	ir D
11	C aa	01 K	ing Ba	10 K	19 98 19 98 10 10	.с. ас Т.	19 g¢ 18 βγ	် ဂို ဂို	7. g.	ម្ ម ម	بع ور 17	ញ ភូព
11/5	ה גד	F 01/6	טמיק	t gt	81/t c ca 0 71/6	19 at I 61/7	t 99 G S1/7	c tt F 41/7	19 gc A 31/8	ic gt V 21/8	19 99 G 11/8	្តី ភ្នំកា
17	ä	T 38	1 .a 5	H H H	9a 92 20 20 20	aa K 21	다. 22 편 다	2 2 2	10 10 10 10 10 10 10 10 10 10 10 10 10 1	25 H	2 C C C	ge ⊢

# SUBSTITUTE SHEET (RULE 26)

## 1761/921  THE CEC THE GAC TAC CTG CCA GCT GAG CAA GCC GAG GCA GAG GAG GAG GAG GAG	O		m D	m u	ru m	
TTC CCC ttc gac tac ctg cca gct gag F P F D Y L P A E 2851/951  Cca gca cga gtg gtg ttc c3g atc tgg P A R V V F Q I W 2941/981  Ccc aag cca gcc aag aca gcc aag aag P K P A K T A K K 3031/1011  aaa aca gtg aag ggc tgg tgg ccc tgt K T V K G W W P C 3121/1041  acc tcc tc ctg tgg tgg ccc tgg C 25 F L W F T S P 2511/1071  acc tcc ttc ctg tgg ttc ccg gcg L F I L L F T S P 3301/1101  Ctc tcc tcc tg tgg gaa ggg gcc gtg L F I L L C C G G 3391/1131  Ctc tcc tgc ctg tgg tcc ctg ggg Cc tgg ttc cct gc gcg ggt Cc tcc tcc tgc tgg gaa ggg gcc gtg L S C P V E G A V 3481/1161  Cc tag gcc tga ttg tcc tgc cag ggt S C P V E G A V 3481/1161  Cc tag gcc tga ttg tcc tgc ccc A S C P V E G D W  361/1221  aaa aaa aaa aaa aaa aa (SEQ ID NO: K K K K K N CSEO ID NO: K			0 4 B	ж т т	က် ကို	00 a 00 a
TTC CCC ttc gac tac ctg cca gcc ccc ttc gac tac ctg cca gcc agc cca gcc agc agc agc agc ag	C33	gard C tgc	gt d A	K K	0 999	Cas 233
TTC CCC ttc gac tac ctg cca gcc ccc ttc gac tac ctg cca gcc agc cca gcc agc agc agc agc ag			tgt C CCt		gtg V ggt G	درد 80:
TTC CCC TCC GAC TAC CTG CCA F P F D Y L P 2851/951 CCa GCA CGA GTG GTG TCC C3G P A R V V F Q 2941/981 GCC AAG CCA GCC AAG ACA GCC P K P A K T A 3031/1011 AAA ACA C GAC GAG GAG A E S E H E E E 3211/1071 T S F L W F T 3301/1101 CCC TCC TCC CTG TGG TTC T S F L W F T 3301/1101 CCC TCC TCC TG TGG TTC T S F L W F T 3301/1101 CCC TCC TCC TG TGG TCC TGC T S F L S C 3781/1131 CCC CAC TCC TGC GTA GAA GGG CCC TCC TCC TGC TGC TGC T S F L L F T 3301/1101 CCC CAC TCC TGC TGC TGC TGC T S F L L F T T S F L W F T T S F L S C T S F L L S C T S F L L S C T S F L L S C T S F L L S C T S C P V E G T S C TCC TGC TGC TGC TGC TGC TGC TGC TGC TGC TGC TCC TGC TGC TGC TGC TGC TGC TGC TGC TCC TGC TGC TGC TGC TGC TGC TGC TGC TGC				s S Ctg	9000 A Cag Q	
2761/921  ttc ccc ttc gac tac ctg F P F D Y L 2851/951  cca gca cga gtg gtg ttc P A R V V F 2941/981  ccc aag cca gcc aag aca P K P A K T 3031/1011  gca gag gtg gag ggc tgg K T V K G W 3121/1071  gca gag gtg gag ctg tt T V K G W 3211/1071  gca gag gt gag ctg tt T S E L W F 3301/1101  ctc tcc ttc ctg tgg ttt L F L L L L 3391/1131  ctc tcc tgc cct gta gaa L S C P V E 3481/1161  ctc tag gcc tga ttg tcc S C C cac ttc cat ctc S C C C C C C C C C C C C C C C C C C C	<b>6</b> 00	cag Q gcc A	t gg K gg gag	a d d d d d d d d d d d d d d d d d d d	999 G G T T G C	1 2 C C C C C C C C C C C C C C C C C C
2761/921  F P F D Y 2851/951  CCa acca cca acc aqu P A R V V 2941/981  CCC aaq cca acc aag P K P A K 3031/1011  aaa aca gtg aag ggc K T V K G 3121/1041  gca gag act gag cat A E S E H 3211/1071  gca gag act ctg tgg T S F L W 3301/1101  Ctc tcc tcc tcc tg tga L S C P V 3481/1161  CCC cac tcc cat cat CC cac tcc cat CC cac tcc cat CC cac tcc cac CC cac tcc cac S A A B 3571/1191  CCC cac tcc cat cat S A A B 361/1221  aaa aaa aaa aaa aa	ctg L	F aca	tgg W gag E	F Ctg L	gaa E E C C	Etc F (SE (SE
2761/921  ttc ccc ttc gac F P F D 2851/951  cca gca cga gtg P A R V 2941/981  ccc aag cca gcc P K P A 3031/1011  aaa aca gtg aag K T V K 3121/1041  gca gag agt gag A E S E 3211/1071  cc tc ttc atc ctg L F L 3301/1101  ctc tcc tcc tcc ctc L S C P 3481/1161  ccc cac ttc cat cc cac tcc cac cc cc cc tcc cc tcc ccc d 3391/1131  ccc cac tcc cac d 361/1221  aaa aaa aaa aaa	tac Y	gtg V A aag K	ggc G Cat	tgg W Ctg L	gta V ttg L	R H aa
2761/921  ttc ccc ttc F P F 2851/951  cca gca cga P A R 2941/981  ccc aag cca P K P 3031/101  aaa aca gtg K T V 3121/104  gca gag agt A E S 3211/107  acc tcc ttc C T S F I I I 3301/110  ctc tcc tcc cgc C C cc tcc C C C C C C C I S A 3571/119  ccc cac ttc B A F I I 3301/110  ctc tcc tcc cgc C C C C C C 3481/116  ccc cac ttc B A F C C C C C C 361/122	Jac	p 5	1 K & a g	ctg 1 ctg L	CCC P 1 1 rga	Cat H H A A A A A
2761 F P 2851 CCC adg P A 3031 P CCC adg P K T 3121 A E 2941 A 12131 190a gag A E 3321 CCC acc CC L S CC L S CC L S CC CCC CCC CCC CCC CCC CCC CCC CCC C	/921 ttc F /951	cga R /981 cca P	9tg V /104 agt S	Etc /110 atc I /113	tgc C /116 /116 gcc A	ttc F /122 aaa K
K A A A C C C C C C C C C C C C C C C C	2761 ccc P P	gca A 2941 aag K K	aca T 3121 gag E	5 3301 ttc F 5391	5 5 3481 tag tag	cac H 3661 aaa K
and the same of th	۳۱ <del>۱۱</del> در		A A A	acc T ctc L	ς τςς 8 τςς	CCC P P R A A A A A A A A A A A A A A A A A
A R R R R R R R R R R R R R R R R R R R	att I	atc A A	cag Q Qta	gac D atc	999a G G G G	s S S S S S S S S S S S S S S S S S S S
F C C C C C C C C C C C C C C C C C C C	n r n	6	gag E gag H t t	P P P P P P P P P P P P P P P P P P P	rga Cag	R B B B B B B B B B B B B B B B B B B B
N S S S S S S S S S S S S S S S S S S S	D	o o o	ក	יו ר א מר יו ר ה	8 8 D T	Z B B B B B B B B B B B B B B B B B B B
F S S C C C C C C C C C C C C C C C C C	£ 20 20 20 20 20 20 20 20 20 20 20 20 20	8 U J	ר ה היים מ ה	2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3	n n	act a
S S C C C C C C C C C C C C C C C C C C	B B C	F I	T C C L	P P P P P P P P P P P P P P P P P P P	n gan	
F F F F F F F F F F F F F F F F F F F	, to	aag c	atg	Jac D B F F		
TE T		() Ph	tt gaa	9 3 3 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	ין דר דר דר	
1/911  99c aac  6 N  1/941  ctg gac  1/1001  tgg ttt  1/1001  ctg gaa  L E  1/1001  ctg gaa  L E  1/101  ctg gaa  L E  1/101  ctg gaa  L E  1/101  ctg gaa  C C  1/1181  ctg ctaa  C C  1/1181  ctg ctaa  C C  1/1181		tg c	.99 t 1031 1031 1061	tt (109)	: tg g (1151 (1151 (ag c	1211 1211
gaa ggc aac E G N 2821/941 agg ctg gac R L D 2911/971 tcc ctg cag S L Q 3001/1001 aag ctg gaa K L E 3091/1031 aag ctg gaa K L E 3181/1061 aag ctg gag K L E 3271/1091 cgg tgg gcc R W A 3361/1121 aag ct gag K L C 3451/1181 aag ct gag C C C C C C C C C C C C C C C C C C C		199 c	194 t 1091/ 199 c	1361/	lag o ( 1 (451/ (60 o (541/	19t to 1631/
				n and and and and and and and and and an	p	1 3 9 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8
4 0 0 0 0 0 0 0 0	ત્ત	υ bn	ט ס	ם ס	ט ט	ע ט
	ກຸ ຄ	ם שם	ה היה היה היה היה היה היה היה היה היה ה	# 1	מ א מ	o e
	ដូ	a a a t	n a		מי ה הי ה מיע מיע	a h
a a a a a a a a a a a a a a a a a a a	ם הי	ag g D at g	a a A	n S t S	a g g Y O	g be
a d a a a a a a a a a a a a a a a a a a	م ب به ن	e g t T D D	a d g d x a D	מ מ ה ה	a S T S	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Dig tip Dig Tr	۵ ۳ ب	0 0 9 2 1 1 1 1	מ מ מ ס מ אים ט	и С Б В В В	ກ ຄຸ ດີ ຄຸດ	ğ th giΩ Di di
THE SERVICE WAS TO THE TO SERVICE TO THE TOTAL	ប័ <del>ភ</del> ភូ	ה ה מית איש	e H Du Ee	8 9 9 9 9 9 1 1	n m n S m m	מים הית צית
29 20 20 20 20 20 20 20 20 20 20 20 20 20	901 3C G( V	cc al 1 961 39 rt	30 Cc. 1021 39 GC 1051	gc ci R 1081 tg at K 1111	ac g A 1141 ca g A 1171	ac ag R 1201 St ti
2701/901  aca gac gcg cat tat cgt tcc ctg gg T D V H Y R S L G 2791/931  tgt acc att gcc aag aag gat gcc tt C T I A K K D A F 2881/961  gac aag trc tcc ttt gat gat ttt ct D V L D D R F L 2971/991  ttg gac cag ctg gat gat gct ttc ca L D Q L D D A F H 3061/1021  gaa gag ggt gag aag aaa ata ctg gc E E G E K K I L A 3151/1051  cag ggc cgg gat gag ccc ac atg aa Q G R D E P N M N 3241/1081  acc atg aag ttc acc ctg tgg cgg cg T M K F I L M R R 3311/1111  ccc acg acc ttc ccg acc tat gct gc T N A F P N Y A A 3421/1141  ccc cca gca tgg gac tgg cct gc tc P P A W D W P A S 3511/1171  acc aca gac aga tgg acc ggc cca cac tc T D R W T G P H S 3601/1201  aac gac aga tgg acc ggc cca cac tc T D R W T G P H S 3601/1201  aac gcc ttt ttg gat cag ctc aga ca	1701/ 103 96 10 1	971/9	tg g D 1061/ 18a g E	tag gr G G 1241/ Icc al	1421/ 1421/ 1421/ 10t of	10a g 1601/ 1601/ 1ac g(

# SUBSTITUTE SHEET (RULE 26)

### SEQUENCE LISTING

<120> DYSFERLIN, A GENE MUTATED IN DISTAL MYOPATHY AND LIMB GIRDLE MUSCULAR DYSTROPHY	
<130> 00786/399WO2	
<150> US 60/097,927 <151> 1998-08-25	
<160> 233	
<170> FastSEQ for Windows Version 3.0	
<210> 1 <211> 6911 <212> DNA <213> Homo sapiens	
<220> <221> CDS <222> (374)(6613)	
<pre>&lt;400&gt; 1 tcgaccgccc agccaggtgc aaaatgccgt gtcattggga gactccgcag ccggagcatt agattacagc tcgacggagc tcgggaaggg cggcgggggt ggaagatgag cagaagcccc tgttctcgga acgccggcg acaagcgggg tgagccgagg cgggggggg acccagcta gcccactgga gcagccgggg gtggcccgtt cccctttaag agcaactgct ctaagccagg agccagagat tcgagccggc ctcgcccagc cagccctctc cagcgagggg acccacaagc ggcgcctcgg ccctcccgac ctttccgagc cctctttgcg ccctgggcgc acggggccct acacgcgcca agc atg ctg agg gtc ttc atc ctc tat gcc gag aac gtc Met Leu Arg Val Phe Ile Leu Tyr Ala Glu Asn Val  10</pre>	60 120 180 240 300 360 409
1 5 10	
cac aca ccc gac acc gac atc agc gat gcc tac tgc tcc gcg gtg ttt  His Thr Pro Asp Thr Asp Ile Ser Asp Ala Tyr Cys Ser Ala Val Phe  15 20 25	457
cac aca ccc gac acc gac atc agc gat gcc tac tgc tcc gcg gtg ttt His Thr Pro Asp Thr Asp Ile Ser Asp Ala Tyr Cys Ser Ala Val Phe	<b>4</b> 57
cac aca ccc gac acc gac atc agc gat gcc tac tgc tcc gcg gtg ttt His Thr Pro Asp Thr Asp Ile Ser Asp Ala Tyr Cys Ser Ala Val Phe 15 20 25  gca ggg gtg aag aag aga acc aaa gtc atc aag aac agc gtg aac cct Ala Gly Val Lys Lys Arg Thr Lys Val Ile Lys Asn Ser Val Asn Pro	
cac aca ccc gac acc gac atc agc gat gcc tac tgc tcc gcg gtg ttt His Thr Pro Asp Thr Asp Ile Ser Asp Ala Tyr Cys Ser Ala Val Phe 15  gca ggg gtg aag aag aga acc aaa gtc atc aag aac agc gtg aac cct Ala Gly Val Lys Lys Arg Thr Lys Val Ile Lys Asn Ser Val Asn Pro 30  gta tgg aat gag gga ttt gaa tgg gac ctc aag ggc atc ccc ctg gac Val Trp Asn Glu Gly Phe Glu Trp Asp Leu Lys Gly Ile Pro Leu Asp	505
cac aca ccc gac acc gac atc agc gat gcc tac tgc tcc gcg gtg ttt His Thr Pro Asp Thr Asp Ile Ser Asp Ala Tyr Cys Ser Ala Val Phe 15  gca ggg gtg aag aag aga acc aaa gtc atc aag aac agc gtg aac cct Ala Gly Val Lys Lys Arg Thr Lys Val Ile Lys Asn Ser Val Asn Pro 30  gta tgg aat gag gga ttt gaa tgg gac ctc aag ggc atc ccc ctg gac Val Trp Asn Glu Gly Phe Glu Trp Asp Leu Lys Gly Ile Pro Leu Asp 45  cag ggc tct gag ctt cat gtg gtg gtc aaa gac cat gag acg atg ggg Gln Gly Ser Glu Leu His Val Val Val Lys Asp His Glu Thr Met Gly	505 553
cac aca ccc gac acc gac atc agc gat gcc tac tgc tcc gcg gtg ttt His Thr Pro Asp Thr Asp Ile Ser Asp Ala Tyr Cys Ser Ala Val Phe 15  gca ggg gtg aag aag aga acc aaa gtc atc aag aac agc gtg aac cct Ala Gly Val Lys Lys Arg Thr Lys Val Ile Lys Asn Ser Val Asn Pro 30  gta tgg aat gag gga ttt gaa tgg gac ctc aag ggc atc ccc ctg gac Val Trp Asn Glu Gly Phe Glu Trp Asp Leu Lys Gly Ile Pro Leu Asp 45  cag ggc tct gag ctt cat gtg gtg gtc aaa gac cat gag acg atg ggg Gln Gly Ser Glu Leu His Val Val Val Lys Asp His Glu Thr Met Gly 65  agg aac agg ttc ctg ggg gaa gcc aag gtc cca ctc cga gag gtc ctc Arg Asn Arg Phe Leu Gly Glu Ala Lys Val Pro Leu Arg Glu Val Leu	505 553 601

ccg Pro 125	ctg Leu	cct Pro	gga Gly	gct Ala	gtg Val 130	ccc Pro	ctg Leu	ttc Phe	ccg Pro	ccc Pro 135	cct Pro	act Thr	cct Pro	ctg Leu	gag Glu 140	793
ccc Pro	tcc Ser	ccg Pro	act Thr	ctg Leu 145	cct Pro	gac Asp	ctg Leu	gat Asp	gta Val 150	gtg Val	gca Ala	gac Asp	aca Thr	gga Gly 155	gga Gly	841
gag Glu	gaa Glu	gac Asp	aca Thr 160	gag Glu	gac Asp	cag Gln	gga Gly	ctc Leu 165	act Thr	gga Gly	gat Asp	gag Glu	gcg Ala 170	gag Glu	cca Pro	889
ttc Phe	ctg Leu	gat Asp 175	caa Gln	agc Ser	gga Gly	ggc Gly	ccg Pro 180	Gly	gct Ala	ccc Pro	acc Thr	acc Thr 185	cca Pro	agg Arg	aaa Lys	937
cta Leu	cct Pro 190	tca Ser	cgt Arg	cct Pro	ccg Pro	ccc Pro 195	cac His	tac Tyr	ccc Pro	Gly ggg	atc Ile 200	aaa Lys	aga Arg	aag Lys	cga Arg	985
agt Ser 205	gcg Ala	cct Pro	aca Thr	tct Ser	aga Arg 210	aag Lys	ctg Leu	ctg Leu	tca Ser	gac Asp 215	aaa Lys	ccg Pro	cag Gln	gat Asp	ttc Phe 220	1033
cag Gln	atc Ile	agg Arg	gtc Val	cag Gln 225	gtg Val	atc Ile	gag Glu	ggg Gly	cgc Arg 230	cag Gln	ctg Leu	ccg Pro	ggg Gly	gtg Val 235	aac Asn	1081
atc Ile	aag Lys	cct Pro	gtg Val 240	gtc Val	aag Lys	gtt Val	acc Thr	gct Ala 245	gca Ala	ggg ggg	cag Gln	acc Thr	aag Lys 250	cgg Arg	acg Thr	1129
cgg <b>Ar</b> g	atc Ile	cac His 255	aag Lys	gga Gly	aac Asn	agc Ser	cca Pro 260	ctc Leu	ttc Phe	aat Asn	gag Glu	act Thr 265	ctt Leu	ttc Phe	ttc Phe	1177
aac Asn	ttg Leu 270	ttt Phe	gac Asp	tct Ser	cct Pro	ggg Gly 275	gag Glu	ctg Leu	ttt Phe	gat Asp	gag Glu 280	ccc Pro	atc Ile	ttt Phe	atc Ile	1225
acg Thr 285	gtg Val	gta Val	gac Asp	tct Ser	cgt Arg 290	tct Ser	ctc Leu	agg Arg	aca Thr	gat Asp 295	gct Ala	ctc Leu	ctc Leu	ggg Gly	gag Glu 300	1273
ttc Phe	cgg Arg	atg Met	gac Asp	gtg Val 305	Gly	Thr	Ile	tac Tyr	Arg	Glu	ccc Pro	cgg Arg	cac His	gcc Ala 315	Tyr	1321
ctc Leu	agg Arg	aag Lys	tgg Trp 320	ctg Leu	ctg Leu	ctc Leu	tca Ser	gac Asp 325	cct Pro	gat Asp	gac Asp	ttc Phe	tct Ser 330	gct Ala	ggg Gly	1369
gcc Ala	aga Arg	ggc Gly 335	tac Tyr	ctg Leu	aaa Lys	aca Thr	agc Ser 340	ctt Leu	tgt Сув	gtg Val	ctg Leu	999 Gly 345	cct Pro	ggg ggg	Asp	1417
gaa Glu	gcg Ala 350	cct Pro	ctg Leu	gag Glu	aga Arg	aaa Lys 355	gac Asp	ccc Pro	tct Ser	gaa Glu	gac Asp 360	aag Lys	gag Glu	gac Asp	att Ile	1465
gaa Glu 365	agc Ser	aac Asn	ctg Leu	ctc Leu	cgg Arg 370	Pro	aca Thr	ggc Gly	gta Val	gcc Ala 375	Leu	cga Arg	gga Gly	gcc Ala	cac His 380	1513
ttc Phe	tgc Cys	ctg Leu	aag Lys	gtc Val 385	Phe	cgg Arg	gcc Ala	gag Glu	gac Asp 390	Leu	ccg Pro	cag Gln	atg Met	gac Asp 395	Asp	1561

gcc Ala	gtg Val	atg Met	gac Asp 400	aac Asn	gtg Val	aaa Lys	cag Gln	atc Ile 405	ttt Phe	ggc Gly	ttc Phe	gag Glu	agt Ser 410	aac Asn	aag Lys	1609
aag Lys	aac Asn	ttg Leu 415	gtg Val	gac Asp	ccc Pro	ttt Phe	gtg Val 420	gag Glu	gtc Val	agc Ser	ttt Phe	gcg Ala 425	ggg Gly	aaa Lys	atg Met	1657
ctg Leu	tgc Cys 430	agc Ser	aag Lys	atc Ile	ttg Leu	gag Glu 435	aag Lys	acg Thr	gcc Ala	aac Asn	cct Pro 440	cag Gln	tgg Trp	aac Asn	cag Gln	1705
aac Asn 445	atc Ile	aca Thr	ctg Leu	cct Pro	gcc Ala 450	atg Met	ttt Phe	ccc Pro	tcc Ser	atg Met 455	tgc Cys	gaa Glu	aaa Lys	atg Met	agg Arg 460	1753
att Ile	cgt Arg	atc Ile	ata Ile	gac Asp 465	tgg Trp	gac Asp	cgc Arg	ctg Leu	act Thr 470	cac His	aat Asn	gac Asp	atc Ile	gtg Val 475	gct Ala	1801
acc Thr	acc Thr	tac Tyr	ctg Leu 480	agt Ser	atg Met	tcg Ser	aaa Lys	atc Ile 485	tct Ser	gcc Ala	cct Pro	gga Gly	gga Gly 490	gaa Glu	ata Ile	1849
gaa Glu	gag Glu	gag Glu 495	cct Pro	gca Ala	ggt Gly	gct Ala	gtc Val 500	aag Lys	cct Pro	tcg Ser	aaa Lys	gcc Ala 505	tca Ser	gac Asp	ttg Leu	1897
gat Asp	gac Asp 510	tac Tyr	ctg Leu	ggc Gly	ttc Phe	ctc Leu 515	ccc Pro	act Thr	ttt Phe	Gly ggg	ccc Pro 520	tgc Cys	tac Tyr	atc Ile	aac Asn	1945
ctc Leu 525	Tyr	ggc Gly	agt Ser	ccc Pro	aga Arg 530	gag Glu	ttc Phe	aca Thr	ggc Gly	ttc Phe 535	cca Pro	gac Asp	ccc Pro	tac Tyr	aca Thr 540	1993
gag Glu	ctc Leu	aac Asn	aca Thr	ggc Gly 545	aag Lys	ggg Gly	gaa Glu	ggt Gly	gtg Val 550	gct Ala	tat Tyr	cgt Arg	ggc Gly	cgg Arg 555	ctt Leu	2041
ctg Leu	ctc Leu	tcc Ser	ctg Leu 560	gag Glu	acc Thr	aag Lys	ctg Leu	gtg Val 565	GIU	cac His	agt Ser	gaa Glu	cag Gln 570	aag Lys	gtg Val	2089
gag Glu	gac Asp	ctt Leu 575	Pro	gcg Ala	gat Asp	gac Asp	atc Ile 580	Leu	cgg Arg	gtg Val	GIU	aag Lys 585	TYL	ctt Leu	agg Arg	2137
ago Aro	g cgc Arg 590	Lys	tac Tyr	tcc	ctg Leu	ttt Phe 595	Ala	gcc Ala	tto Phe	tac Tyr	tca Ser 600	WIG	acc Thr	atg Met	ctg Leu	2185
caq Glr 605	Asp	gto Val	gat Asp	gat Asp	gcc Ala 610	Ile	cag Gln	ttt Phe	gag Glu	gto Val 615	. ser	atc : Ile	ggg Gly	aac Asn	tac Tyr 620	2233
ggg Gl	y aac y Asi	aaq Lys	g tto 3 Phe	gac Asp 625	Met	acc Thr	tgc Cys	cto Lev	g ccc Pro 630	Let	g gcc Ala	tcc Ser	acc Thr	act Thr 635	cag Gln	2281
tac Ty:	c ago r Sei	c cgt	g Ala 640	. Val	ttt Phe	gac Asp	ggg Gly	g tgo 7 Cys 645	3 Hls	tac Tyi	tac Tyi	tac Tyr	cta Lev 650	Pro	tgg Trp	2329

ggt Gly	aac Asn	gtg Val 655	aaa Lys	cct Pro	gtg Val	gtg Val	gtg Val 660	ctg Leu	tca Ser	tcc Ser	tac Tyr	tgg Trp 665	gag Glu	gac Asp	atc Ile	2377
agc Ser	cat His 670	aga Arg	atc Ile	gag Glu	act Thr	cag Gln 675	aac Asn	cag Gln	ctg Leu	ctt Leu	680 61y	att Ile	gct Ala	gac Asp	cgg Arg	2425
ctg Leu 685	gaa Glu	gct Ala	Gly ggc	ctg Leu	gag Glu 690	cag Gln	gtc Val	cac His	ctg Leu	gcc Ala 695	ctg Leu	aag Lys	gcg Ala	cag Gln	tgc Cys 700	2473
tcc Ser	acg Thr	gag Glu	gac Asp	gtg Val 705	gac Asp	tcg Ser	ctg Leu	gtg Val	gct Ala 710	cag Gln	ctg Leu	acg Thr	gat Asp	gag Glu 715	ctc Leu	2521
atc Ile	gca Ala	ggc Gly	tgc Cys 720	agc Ser	cag Gln	cct Pro	ctg Leu	ggt Gly 725	gac Asp	atc Ile	cat His	gag Glu	aca Thr 730	ccc Pro	tct Ser	2569
gcc Ala	acc Thr	cac His 735	ctg Leu	gac Asp	cag Gln	tac Tyr	ctg Leu 740	tac Tyr	cag Gln	ctg Leu	cgc Arg	acc Thr 745	cat His	cac His	ctg Leu	2617
agc Ser	caa Gln 750	atc Ile	act Thr	gag Glu	gct Ala	gcc Ala 755	ctg Leu	gcc Ala	ctg Leu	aag Lys	ctc Leu 760	ggc Gly	cac His	agt Ser	gag Glu	2665
ctc Leu 765	cct Pro	gca Ala	gct Ala	ctg Leu	gag Glu 770	cag Gln	gcg Ala	gag Glu	gac Asp	tgg Trp 775	ctc Leu	ctg Leu	egt Arg	ctg Leu	cgt Arg 780	2713
gcc Ala	ctg Leu	gca Ala	gag Glu	gag Glu 785	ccc Pro	cag Gln	aac Asn	agc Ser	ctg Leu 790	ccg Pro	gac Asp	atc Ile	gtc Val	atc Ile 795	tgg Trp	2761
atg Met	ctg Leu	cag Gln	gga Gly 800	gac Asp	aag Lys	cgt Arg	gtg Val	gca Ala 805	tac Tyr	cag Gln	cgg Arg	gtg Val	ccc Pro 810	gcc Ala	cac His	2809
caa Gln	gtc Val	ctc Leu 815	Phe	tcc Ser	cgg Arg	cgg Arg	ggt Gly 820	gcc Ala	aac Asn	tac Tyr	tgt Cys	ggc Gly 825	aag Lys	aat Asn	tgt Cys	2857
ggg Gly	aag Lys 830	cta Leu	cag Gln	aca Thr	atc Ile	ttt Phe 835	ctg Leu	aaa Lys	tat Tyr	ccg Pro	atg Met 840	gag Glu	aag Lys	gtg Val	cct Pro	2905
ggc Gly 845	Ala	cgg Arg	atg Met	cca Pro	gtg Val 850	cag Gln	ata Ile	cgg Arg	gtc Val	aag Lys 855	ctg Leu	tgg Trp	ttt Phe	ggg Gly	ctc Leu 860	2953
tct Ser	gtg Val	gat Asp	gag Glu	aag Lys 865	Glu	ttc Phe	aac Asn	cag Gln	ttt Phe 870	gct Ala	gag Glu	ggg	aag Lys	ctg Leu 875	tct Ser	3001
gtc Val	ttt Phe	gct Ala	gaa Glu 880	Thr	tat Tyr	gag Glu	aac Asn	gag Glu 885	Thr	aag Lys	ttg Leu	gcc Ala	ctt Leu 890	vai	ggg ggg	3049
aac Asn	tgg Trp	ggc Gly 895	Thr	acg Thr	ggc Gly	ctc Leu	acc Thr 900	Tyr	ccc	aag Lys	ttt Phe	tct Ser 905	Asp	gto Val	acg Thr	3097
ggc Gly	aag Lys 910	Ile	aag Lys	cta Leu	ccc Pro	aag Lys 915	Asp	agc Ser	ttc Phe	cgc Arg	ccc Pro 920	Ser	gcc Ala	ggc	tgg Trp	3145

WO 00/11157 PCT/US99/19395

acc Thr 925	tgg Trp	gct Ala	gga Gly	gat Asp	tgg Trp 930	ttc Phe	gtg Val	tgt Cys	ccg Pro	gag Glu 935	aag Lys	act Thr	ctg Leu	ctc Leu	cat His 940	3193
gac Asp	atg Met	gac Asp	gcc Ala	ggt Gly 945	cac His	ctg Leu	agc Ser	ttc Phe	gtg Val 950	gaa Glu	gag Glu	gtg Val	ttt Phe	gag Glu 955	aac Asn	3241
cag Gln	acc Thr	cgg Arg	ctt Leu 960	ccc Pro	gga Gly	ggc Gly	cag Gln	tgg Trp 965	atc Ile	tac Tyr	atg Met	agt Ser	gac Asp 970	aac Asn	tac Tyr	3289
acc Thr	gat Asp	gtg Val 975	aac Asn	ggg Gly	gag Glu	aag Lys	gtg Val 980	ctt Leu	ccc Pro	aag Lys	gat Asp	gac Asp 985	att Ile	gag Glu	tgc Cys	3337
cca Pro	ctg Leu 990	Gly	tgg Trp	aag Lys	tgg Trp	gaa Glu 995	Asp	gag Glu	gaa Glu	tgg Trp	tcc Ser 1000	Thr	gac Asp	ctc Leu	aac Asn	3385
cgg Arg 100!	Ala	gtc Val	gat Asp	gag Glu	caa Gln 1010	Gly	tgg Trp	gag Glu	tat Tyr	agc Ser 101!	тте	acc Thr	atc Ile	ccc Pro	ccg Pro 1020	3433
gag Glu	cgg Arg	aag Lys	ccg Pro	aag Lys 102!	cac His	tgg Trp	gtc Val	cct Pro	gct Ala 1030	GLu	aag Lys	atg Met	tac Tyr	tac Tyr 103	Thr	3481
cac His	cga Arg	cgg Arg	cgg Arg 1040	Arg	tgg Trp	gtg Val	cgc Arg	ctg Leu 104!	Arg	agg Arg	agg Arg	gat Asp	ctc Leu 1050	Ser	caa Gln	3529
atg Met	gaa Glu	gca Ala 105	Leu	aaa Lys	agg Arg	cac His	agg Arg 106	Gln	gcg Ala	gag Glu	gcg Ala	gag Glu 106	GIY	gag Glu	ggc Gly	3577
tgg Trp	gag Glu 107	Tyr	gcc Ala	tct Ser	ctt Leu	ttt Phe 107	Gly	tgg Trp	aag Lys	ttc Phe	cac His 1080	Leu	gag Glu	tac Tyr	cgc Arg	3625
aag Lys 108	Thr	gat Asp	gcc Ala	ttc Phe	cgc Arg 109	Arg	cgc Arg	cgc Arg	tgg Trp	cgc Arg 109	Arg	cgc Arg	atg Met	gag Glu	cca Pro 1100	3673
ctg Leu	gag Glu	aag Lys	Thr	Glv	cct Pro 5	Ala	Ala	Val	Phe	Ala	Leu	Glu	Gly	gcc Ala 111	Leu	3721
ggc	ggc Gly	gtg Val	atg Met 112	Asp	gac Asp	aag Lys	agt Ser	gaa Glu 112	Asp	tcc Ser	atg Met	tcc Ser	gtc Val 113	Ser	acc Thr	3769
ttg Leu									att	tcc	tac	ata	ttc	gac	tat	3817
	agc Ser	ttc Phe 113	Gly	gtg Val	aac Asn	aga Arg	Pro 114	Thr	Ile	Ser	Cys	Ile 114	Pne	Āsp	Tyr	
Gly ggg	Ser	Phe 113 cgc Arg	Gly 5	Val cat	Asn cta	Arg	Pro 114 tgc Cys	Thr O tac	Ile atg	ser	cag	11e 114 gcc Ala	Pne 5 cgg	gac	Tyr ctg Leu	3865

ttc ctg cac cag Phe Leu His Gln	agc cag aag Ser Gln Lys 1185	acg gtg gtg Thr Val Val 1190	Val Lys Asn	acc ctt aac Thr Leu Asn 1195	3961
ccc acc tgg gac Pro Thr Trp Asp 1200	Gln Thr Leu	atc ttc tac Ile Phe Tyr 1205	Glu Ile Glu	atc ttt ggc Ile Phe Gly 1210	4009
gag ccg gcc aca Glu Pro Ala Thr 1215	gtt gct gag Val Ala Glu	caa ccg ccc Gln Pro Pro 1220	agc att gtg Ser Ile Val 1225	gtg gag ctg Val Glu Leu	4057
tac gac cat gac Tyr Asp His Asp 1230	act tat ggt Thr Tyr Gly 1235	Ala Asp Glu	ttt atg ggt Phe Met Gly 1240	cgc tgc atc Arg Cys Ile	4105
tgt caa ccg agt Cys Gln Pro Ser 1245	ctg gaa cgg Leu Glu Arg 1250	atg cca cgg Met Pro Arg	ctg gcc tgg Leu Ala Trp 1255	ttc cca ctg Phe Pro Leu 1260	4153
acg agg ggc agc Thr Arg Gly Ser	cag ccg tcg Gln Pro Ser 1265	ggg gag ctg Gly Glu Leu 127	Leu Ala Ser	ttt gag ctc Phe Glu Leu 1275	4201
atc cag aga gag Ile Gln Arg Glu 1280	Lys Pro Ala	atc cac cat Ile His His 1285	Ile Pro Gly	ttt gag gtg Phe Glu Val 1290	4249
cag gag aca tca Gln Glu Thr Ser 1295	agg atc ctg Arg Ile Leu	gat gag tct Asp Glu Ser 1300	gag gac aca Glu Asp Thr 1305	Asp Leu Pro	4297
tac cca cca ccc Tyr Pro Pro Pro 1310	cag agg gag Gln Arg Glu 1315	Ala Asn Ile	tac atg gtt Tyr Met Val 1320	cct cag aac Pro Gln Asn	4345
atc aag cca gcg Ile Lys Pro Ala 1325	ctc cag cgt Leu Gln Arg 1330	acc gcc atc Thr Ala Ile	gag atc ctg Glu Ile Leu 1335	gca tgg ggc Ala Trp Gly 1340	4393
ctg cgg aac atg Leu Arg Asn Met	aag agt tac Lys Ser Tyr 1345	cag ctg gcc Gln Leu Ala 135	Asn Ile Ser	tcc ccc agc Ser Pro Ser 1355	4441
ctc gtg gta gag Leu Val Val Glu 1360	Cys Gly Gly	cag acg gtg Gln Thr Val 1365	cag tcc tgt Gln Ser Cys	gtc atc agg Val Ile Arg 1370	4489
aac ctc cgg aag Asn Leu Arg Lys 1375	aac ccc aac Asn Pro Asn	ttt gac atc Phe Asp Ile 1380	tgc acc ctc Cys Thr Leu 1385	Phe Met Glu	4537
gtg atg ctg ccc Val Met Leu Pro 1390	agg gag gag Arg Glu Glu 1395	Leu Tyr Cys	ccc ccc atc Pro Pro Ile 1400	acc gtc aag Thr Val Lys	4585
gtc atc gat aac Val Ile Asp Asn 1405	cgc cag ttt Arg Gln Phe 1410	ggc cgc cgg Gly Arg Arg	cct gtg gtg Pro Val Val 1415	ggc cag tgt Gly Gln Cys 1420	4633
acc atc cgc tcc Thr Ile Arg Ser	ctg gag agc Leu Glu Ser 1425	ttc ctg tgt Phe Leu Cys 143	Asp Pro Tyr	tcg gcg gag Ser Ala Glu 1435	4681
agt cca tcc cca Ser Pro Ser Pro 1440	Gln Gly Gly	cca gac gat Pro Asp Asp 1445	gtg agc cta Val Ser Leu	ctc agt cct Leu Ser Pro 1450	4729

eja aaa	gaa Glu	gac Asp 1455	Val	ctc Leu	atc Ile	gac Asp	att Ile 1460	Asp	gac Asp	aag Lys	gag Glu	ccc Pro 1465		atc Ile	ccc Pro	4777
atc Ile	cag Gln 1470	gag Glu )	gaa Glu	gag Glu	ttc Phe	atc Ile 1475	Asp	tgg Trp	tgg Trp	agc Ser	aaa Lys 1480	FILE	ttt Phe	gcc Ala	tcc Ser	4825
ata Ile 1485	Gly	gag Glu	agg Arg	gaa Glu	aag Lys 1490	Сув	ggc Gly	tcc Ser	tac Tyr	ctg Leu 1495	GIU	aag Lys	gat Asp	ttt Phe	gac Asp 1500	4873
acc Thr	ctg Leu	aag Lys	gtc Val	tat Tyr 1505	Asp	aca Thr	cag Gln	ctg Leu	gag Glu 1510	ABII	gtg Val	gag Glu	gcc Ala	ttt Phe 1519	014	4921
ggc Gly	ctg Leu	tct Ser	gac Asp 1520	Phe	tgt Cys	aac Asn	acc Thr	ttc Phe 1525	rÀa	ctg Leu	tac Tyr	cgg Arg	ggc Gly 1530	гдэ	acg Thr	4969
cag Gln	gag Glu	gag Glu 1535	Thr	gaa Glu	gat Asp	cca Pro	tct Ser 1540	vaı	att Ile	ggt Gly	gaa Glu	ttt Phe 1545	цyы	ggc Gly	ctc Leu	5017
ttc Phe	aaa Lys 155	att Ile O	tat Tyr	ccc Pro	ctc Leu	cca Pro 155	Glu	gac Asp	cca Pro	gcc Ala	atc Ile 1560	Pro	atg Met	ccc Pro	cca Pro	5065
aga Arg 156	Gln	ttc Phe	cac His	cag Gln	ctg Leu 157	Ala	gcc Ala	cag Gln	gga Gly	ccc Pro 157	GIII	gag Glu	tgc Cys	ttg Leu	gtc Val 1580	5113
cgt Arg	atc Ile	tac Tyr	att Ile	gtc Val 158	Arg	gca Ala	ttt Phe	ggc	ctg Leu 159	GIn	ccc Pro	aag Lys	gac Asp	ccc Pro 159	Maii	5161
gga Gly	aag Lys	tgt Cys	gat Asp 160	Pro	tac Tyr	atc Ile	aag Lys	atc Ile 160	ser	ata Ile	Gly	aag Lys	aaa Lys 161	Ser	gtg Val	5209
agt Ser	gac Asp	cag Gln 161	Asp	aac Asn	tac Tyr	atc Ile	ccc Pro 162	Cys	acg Thr	ctg Leu	gag Glu	ccc Pro 162	var	ttt Phe	gga Gly	5257
aag Lys	atg Met 163	ttc Phe	gag Glu	ctg Leu	acc Thr	tgc Cys 163	rnr	Leu	cct Pro	Leu	GIU	ъур	gac Asp	cta Leu	aag Lys	5305
ato Ile 164	Thr	ctc Leu	tat Tyr	gac Asp	tat Tyr 165	Asp	ctc Leu	cto Leu	tcc Ser	aag Lys 165	Asp	gaa Glu	aag Lys	atc Ile	ggt Gly 1660	5353
gag Glu	g acc	gtc Val	gtc Val	gac Asp 166	Leu	gag Glu	aac Asn	agg Arg	ctg Leu 167	Leu	tcc Ser	aag Lys	ttt Phe	ggg Gly 167	ALG	5401
cgo <b>A</b> ro	tgt g Cys	gga Gly	cto Leu 168	Pro	caç Glr	acc Thr	tac Tyr	tgt Cys 168	yaı	tct Ser	gga Gly	ccg Pro	aac Asn 169	i GTI	tgg Trp	5449
cgg	g gad g Asj	caç Glr 169	1 Leu	cgc Arg	e ccc	tco Ser	caç Glr 170	ı Let	c cto 1 Leu	cac His	cto Lev	tto Phe 170	Cys	caq Glr	g cag n Gln	5497

cat His	aga gto Arg Val 1710	aag Lys	gca Ala	cct Pro	gtg Val 1715	Tyr	cgg Arg	aca Thr	gac Asp	egt Arg 1720	vaı	atg Met	ttt Phe	cag Gln	5545
gat Asp 172	aaa gaa Lys Glu 5	tat Tyr	tcc Ser	att Ile 1730	Glu	gag Glu	ata Ile	gag Glu	gct Ala 1735	GTĀ	agg Arg	atc Ile	cca Pro	aac Asn 1740	5593
cca Pro	cac cto	ggc Gly	cca Pro 174	Val	gag Glu	gag Glu	cgt Arg	ctg Leu 1750	Ala	ctg Leu	cat His	gtg Val	ctt Leu 1759	GIN	5641
cag Gln	cag ggo Gln Gl	ctg Leu 176	Val	ccg Pro	gag Glu	cac His	gtg Val 1765	Glu	tca Ser	cgg Arg	ccc Pro	ctc Leu 1770	Tyr	agc Ser	5689
ccc Pro	ctg cac Leu Gli	) Pro	gac Asp	atc Ile	gag Glu	cag Gln 1780	GIA	aag Lys	ctg Leu	cag Gln	atg Met 178!	Trp	gtc Val	gac Asp	5737
cta Leu	ttt ccc Phe Pro 1790	g aag o Lys	gcc Ala	ctg Leu	ggg Gly 1795	Arg	cct Pro	gga Gly	cct Pro	ccc Pro 1800	Phe	aac Asn	atc Ile	acc Thr	5785
cca Pro 180	cgg ag Arg Arg	a gcc g Ala	aga Arg	agg Arg 1810	Phe	ttc Phe	ctg Leu	cgt Arg	tgt Cys 181	11e	atc Ile	tgg Trp	aat Asn	acc Thr 1820	5833
aga Arg	gat gto	g atc l Ile	ctg Leu 182	Asp	gac Asp	ctg Leu	agc Ser	ctc Leu 1830	Thr	ggg Gly	gag Glu	aag Lys	atg Met 183	ser	5881
gac Asp	att ta Ile Ty	t gtg r Val 184	. Lys	ggt Gly	tgg Trp	atg Met	att Ile 184	GIA	ttt Phe	gaa Glu	gaa Glu	cac His 185	гÃа	caa Gln	5929
aag Lys	aca ga Thr As 18	p Val	cat His	tat Tyr	cgt Arg	tcc Ser 186	Leu	gga Gly	ggt Gly	gaa Glu	ggc Gly 186	Asn	ttc Phe	aac Asn	5977
tgg Trp	agg tt Arg Ph 1870	c att e Ile	ttc Phe	ccc Pro	ttc Phe 187	Asp	tac Tyr	ctg Leu	cca Pro	gct Ala 188	Glu	caa Gln	gtc Val	tgt Cys	6025
acc Thr 188	att gc lle Al	c aaq a Lys	g aag B Lys	gat Asp 189	Ala	ttc Phe	tgg Trp	Arg	ctg Leu 189	Asp	aag Lys	act Thr	gag Glu	agc Ser 1900	6073
aaa Lys	atc cc Ile Pr	a gca o Ala	a cga a Arg 190	Val	gtg Val	ttc Phe	cag Gln	atc Ile 191	Trp	gac Asp	aat Asn	gac Asp	aag Lys 191	Phe	6121
tco Ser	ttt ga	t gat p Asj 192	o Ph€	ctg Leu	ggc Gly	tcc Ser	ctg Leu 192	Gln	ctc Leu	gat Asp	cto Leu	aac Asn 193	Arg	atg Met	6169
Pro	aag co Lys Pr 19	a gco o Ala	c aaç a Lys	aca Thr	gcc Ala	aag Lys 194	Lys	tgc Cya	tcc Ser	ttg Leu	gac Asr 194	GIN	ctg Leu	gat Asp	6217
gat Asj	gct tt Ala Ph 1950	c ca e Hi	c cca s Pro	gaa Glu	tgg Trp 195	Phe	gtç Val	tcc Ser	ctt Leu	ttt Phe 196	GIU	g cag Gln	g aaa 1 Lys	a aca 3 Thr	6265
gt: Va. 19	g aag go L Lys G 55	jc tg .y Tr	g tg p Tr	g ccc Pro 197	су Сув	gta Val	a gca L Ala	a gaa a Glu	gag Glu 197	ı GIY	gaç Glu	g aag 1 Lys	g aaa B Lys	ata Ile 1980	6313

ctg Leu	gcg Ala	ggc Gly	aag Lys	ctg Leu 1985	Glu	atg Met	acc Thr	ttg Leu	gag Glu 1990	TTG	gta Val	gca Ala	gag Glu	agt Ser 1995	GIU	6361
cat His	gag Glu	gag Glu	cgg Arg 2000	Pro	gct Ala	ggc Gly	cag Gln	ggc Gly 2005	Arg	gat Asp	gag Glu	ccc Pro	aac Asn 2010	atg Met )	aac Asn	6409
cct Pro	aag Lys	ctt Leu 201	Glu	gac Asp	cca Pro	agg Arg	cgc Arg 2020	Pro	gac Asp	acc Thr	tcc Ser	ttc Phe 2025	Leu	tgg Trp	ttt Phe	6457
acc Thr	tcc Ser 2030	Pro	tac Tyr	aag Lys	acc Thr	atg Met 203	Lys	ttc Phe	atc Ile	ctg Leu	tgg Trp 2040	Arg	cgt Arg	ttc Phe	cgg Arg	6505
tgg Trp 204!	Ala	atc Ile	atc Ile	ctc Leu	ttc Phe 2050	Ile	atc Ile	ctc Leu	ttc Phe	atc Ile 205!	Leu	ctg Leu	ctg Leu	ttc Phe	ctg Leu 2060	6553
gcc Ala	atc Ile	ttc Phe	atc Ile	tac Tyr 206	Ala	ttc Phe	ccg Pro	aac Asn	tat Tyr 207	Ala	gcc Ala	atg Met	aag Lys	ctg Leu 207	Vai	6601
aag Lys	ccc Pro	ttc Phe	agc Ser 208		ggac	tct (	cctg	ccct	gt a	gaag	gggc	c gt	<b>3</b> 333	tccc		6653
agg gtt acg	cctg	att aca ttt	gtcc tgga ggat	tgcc gctc cagc	ag g tg a tc a	gtgg	gcag.	a cad	gaca ttcc	gatg atca	gac	cggc	ctc	CCCC	cctcct cccaga aaccca aaaaaa	6713 6773 6833 6893 6911

<210> 2 <211> 2080 <212> PRT <213> Homo sapiens

<400> 2 Met Leu Arg Val Phe Ile Leu Tyr Ala Glu Asn Val His Thr Pro Asp 10 Thr Asp Ile Ser Asp Ala Tyr Cys Ser Ala Val Phe Ala Gly Val Lys 20 25 Lys Arg Thr Lys Val Ile Lys Asn Ser Val Asn Pro Val Trp Asn Glu 45 40 35 Gly Phe Glu Trp Asp Leu Lys Gly Ile Pro Leu Asp Gln Gly Ser Glu 55 Leu His Val Val Lys Asp His Glu Thr Met Gly Arg Asn Arg Phe 75 70 Leu Gly Glu Ala Lys Val Pro Leu Arg Glu Val Leu Ala Thr Pro Ser 90 Leu Ser Ala Ser Phe Asn Ala Pro Leu Leu Asp Thr Lys Lys Gln Pro 110 105 100 Thr Gly Ala Ser Leu Val Leu Gln Val Ser Tyr Thr Pro Leu Pro Gly 120 125 115 Ala Val Pro Leu Phe Pro Pro Pro Thr Pro Leu Glu Pro Ser Pro Thr 135 140 Leu Pro Asp Leu Asp Val Val Ala Asp Thr Gly Gly Glu Glu Asp Thr 155 150 Glu Asp Gln Gly Leu Thr Gly Asp Glu Ala Glu Pro Phe Leu Asp Gln 170 175 165 Ser Gly Gly Pro Gly Ala Pro Thr Thr Pro Arg Lys Leu Pro Ser Arg 190 185 180 Pro Pro Pro His Tyr Pro Gly Ile Lys Arg Lys Arg Ser Ala Pro Thr 205 200 195

	210		Leu			215					220				
Gln 225	Val	Ile	Glu	Gly	Arg 230	Gln	Leu	Pro	Gly	Val 235	Asn	Ile	Lys	Pro	<b>Val</b> 240
Val	ГÀв	Val	Thr	Ala 245	Ala	Gly	Gln	Thr	Lys 250	Arg	Thr	Arg	Ile	His 255	ГÀв
Gly	Asn	Ser	Pro 260	Leu	Phe	Asn	Glu	Thr 265	Leu	Phe	Phe	Asn	Leu 270	Phe	Asp
Ser	Pro	Gly 275	Glu	Leu	Phe	Asp	Glu 280	Pro	Ile	Phe	Ile	Thr 285	Val	Val	Asp
Ser	Arg 290	Ser	Leu	Arg	Thr	Asp 295	Ala	Leu	Leu	Gly	Glu 300	Phe	Arg	Met	Asp
305	Gly		Ile		310					315					320
Leu			Ser	325					330					335	
	_		Ser 340					345					350		
	_	355	Asp				360					365			
	370		Thr			375					380				
385			Ala		390					395					400
		_	Gln	405					410					415	
-			Val 420					425					430		
		435	Lys				440					445			
	450		Phe			455					460				
465			Arg		470					475					480
			rys	485					490					495	
	_		Val 500					505					510		
_		515	Pro				520					525			
	530		Phe			535					540				
545	_		Glu		550					555					560
		_	Leu	565					570					575	
			Ile 580					585					590		
		595	Ala				600					605			
_	610		Gln			615					620				
625			Сув		630					635					640
			Gly	645					650					655	
			Val 660					665					670		
		675	Asn				680					685			
	690		Val			695	,				700				
705			Leu		710					715					720
Ser	Gln	Pro	Leu	Gly 725		Ile	His	Glu	730	Pro	Ser	. Ala	Thr	735	Leu

Asp Gln Tyr Leu Tyr Gln Leu Arg Thr His His Leu Ser Gln Ile Thr Glu Ala Ala Leu Ala Leu Lys Leu Gly His Ser Glu Leu Pro Ala Ala Leu Glu Gln Ala Glu Asp Trp Leu Leu Arg Leu Arg Ala Leu Ala Glu Glu Pro Gln Asn Ser Leu Pro Asp Ile Val Ile Trp Met Leu Gln Gly Asp Lys Arg Val Ala Tyr Gln Arg Val Pro Ala His Gln Val Leu Phe Ser Arg Arg Gly Ala Asn Tyr Cys Gly Lys Asn Cys Gly Lys Leu Gln Thr Ile Phe Leu Lys Tyr Pro Met Glu Lys Val Pro Gly Ala Arg Met Pro Val Gln Ile Arg Val Lys Leu Trp Phe Gly Leu Ser Val Asp Glu Lys Glu Phe Asn Gln Phe Ala Glu Gly Lys Leu Ser Val Phe Ala Glu Thr Tyr Glu Asn Glu Thr Lys Leu Ala Leu Val Gly Asn Trp Gly Thr Thr Gly Leu Thr Tyr Pro Lys Phe Ser Asp Val Thr Gly Lys Ile Lys Leu Pro Lys Asp Ser Phe Arg Pro Ser Ala Gly Trp Thr Trp Ala Gly Asp Trp Phe Val Cys Pro Glu Lys Thr Leu Leu His Asp Met Asp Ala Gly His Leu Ser Phe Val Glu Glu Val Phe Glu Asn Gln Thr Arg Leu Pro Gly Gly Gln Trp Ile Tyr Met Ser Asp Asn Tyr Thr Asp Val Asn Gly Glu Lys Val Leu Pro Lys Asp Asp Ile Glu Cys Pro Leu Gly Trp Lys Trp Glu Asp Glu Glu Trp Ser Thr Asp Leu Asn Arg Ala Val Asp Glu Gln Gly Trp Glu Tyr Ser Ile Thr Ile Pro Pro Glu Arg Lys Pro Lys His Trp Val Pro Ala Glu Lys Met Tyr Tyr Thr His Arg Arg 1025 1030 1035 104 Arg Trp Val Arg Leu Arg Arg Arg Asp Leu Ser Gln Met Glu Ala Leu Lys Arg His Arg Gln Ala Glu Ala Glu Gly Glu Gly Trp Glu Tyr Ala Ser Leu Phe Gly Trp Lys Phe His Leu Glu Tyr Arg Lys Thr Asp Ala Phe Arg Arg Arg Trp Arg Arg Met Glu Pro Leu Glu Lys Thr Gly Pro Ala Ala Val Phe Ala Leu Glu Gly Ala Leu Gly Gly Val Met Asp Asp Lys Ser Glu Asp Ser Met Ser Val Ser Thr Leu Ser Phe Gly Val Asn Arg Pro Thr Ile Ser Cys Ile Phe Asp Tyr Gly Asn Arg Tyr His Leu Arg Cys Tyr Met Tyr Gln Ala Arg Asp Leu Ala Ala Met Asp Lys Asp Ser Phe Ser Asp Pro Tyr Ala Ile Val Ser Phe Leu His Gln Ser Gln Lys Thr Val Val Val Lys Asn Thr Leu Asn Pro Thr Trp Asp Gln Thr Leu Ile Phe Tyr Glu Ile Glu Ile Phe Gly Glu Pro Ala Thr Val Ala Glu Gln Pro Pro Ser Ile Val Val Glu Leu Tyr Asp His Asp Thr Tyr Gly Ala Asp Glu Phe Met Gly Arg Cys Ile Cys Gln Pro Ser Leu Glu Arg Met Pro Arg Leu Ala Trp Phe Pro Leu Thr Arg Gly Ser 

Gln Pro Ser Gly Glu Leu Leu Ala Ser Phe Glu Leu Ile Gln Arg Glu 1270 · 1275 Lys Pro Ala Ile His His Ile Pro Gly Phe Glu Val Gln Glu Thr Ser Arg Ile Leu Asp Glu Ser Glu Asp Thr Asp Leu Pro Tyr Pro Pro Pro Gln Arg Glu Ala Asn Ile Tyr Met Val Pro Gln Asn Ile Lys Pro Ala Leu Gln Arg Thr Ala Ile Glu Ile Leu Ala Trp Gly Leu Arg Asn Met Lys Ser Tyr Gln Leu Ala Asn Ile Ser Ser Pro Ser Leu Val Val Glu Cys Gly Gly Gln Thr Val Gln Ser Cys Val Ile Arg Asn Leu Arg Lys Asn Pro Asn Phe Asp Ile Cys Thr Leu Phe Met Glu Val Met Leu Pro Arg Glu Glu Leu Tyr Cys Pro Pro Ile Thr Val Lys Val Ile Asp Asn Arg Gln Phe Gly Arg Arg Pro Val Val Gly Gln Cys Thr Ile Arg Ser Leu Glu Ser Phe Leu Cys Asp Pro Tyr Ser Ala Glu Ser Pro Ser Pro 1425 Gln Gly Gly Pro Asp Asp Val Ser Leu Leu Ser Pro Gly Glu Asp Val Leu Ile Asp Ile Asp Asp Lys Glu Pro Leu Ile Pro Ile Gln Glu Glu Glu Phe Ile Asp Trp Trp Ser Lys Phe Phe Ala Ser Ile Gly Glu Arg Glu Lys Cys Gly Ser Tyr Leu Glu Lys Asp Phe Asp Thr Leu Lys Val Tyr Asp Thr Gln Leu Glu Asn Val Glu Ala Phe Glu Gly Leu Ser Asp Phe Cys Asn Thr Phe Lys Leu Tyr Arg Gly Lys Thr Gln Glu Glu Thr Glu Asp Pro Ser Val Ile Gly Glu Phe Lys Gly Leu Phe Lys Ile Tyr Pro Leu Pro Glu Asp Pro Ala Ile Pro Met Pro Pro Arg Gln Phe His Gln Leu Ala Ala Gln Gly Pro Gln Glu Cys Leu Val Arg Ile Tyr Ile Val Arg Ala Phe Gly Leu Gln Pro Lys Asp Pro Asn Gly Lys Cys Asp 1590 1595 Pro Tyr Ile Lys Ile Ser Ile Gly Lys Lys Ser Val Ser Asp Gln Asp Asn Tyr Ile Pro Cys Thr Leu Glu Pro Val Phe Gly Lys Met Phe Glu Leu Thr Cys Thr Leu Pro Leu Glu Lys Asp Leu Lys Ile Thr Leu Tyr Asp Tyr Asp Leu Leu Ser Lys Asp Glu Lys Ile Gly Glu Thr Val Val Asp Leu Glu Asn Arg Leu Leu Ser Lys Phe Gly Ala Arg Cys Gly Leu Pro Gln Thr Tyr Cys Val Ser Gly Pro Asn Gln Trp Arg Asp Gln Leu Arg Pro Ser Gln Leu Leu His Leu Phe Cys Gln Gln His Arg Val Lys Ala Pro Val Tyr Arg Thr Asp Arg Val Met Phe Gln Asp Lys Glu Tyr Ser Ile Glu Glu Ile Glu Ala Gly Arg Ile Pro Asn Pro His Leu Gly Pro Val Glu Glu Arg Leu Ala Leu His Val Leu Gln Gln Gln Leu Val Pro Glu His Val Glu Ser Arg Pro Leu Tyr Ser Pro Leu Gln Pro Asp Ile Glu Gln Gly Lys Leu Gln Met Trp Val Asp Leu Phe Pro Lys 

```
Ala Leu Gly Arg Pro Gly Pro Pro Phe Asn Ile Thr Pro Arg Arg Ala
                                                1805
                        . 1800
        1795
Arg Arg Phe Phe Leu Arg Cys Ile Ile Trp Asn Thr Arg Asp Val Ile
                                            1820
                        1815
    1810
Leu Asp Asp Leu Ser Leu Thr Gly Glu Lys Met Ser Asp Ile Tyr Val
                                        1835
                    1830
Lys Gly Trp Met Ile Gly Phe Glu Glu His Lys Gln Lys Thr Asp Val
                                                        1855
                                    1850
                1845
His Tyr Arg Ser Leu Gly Gly Glu Gly Asn Phe Asn Trp Arg Phe Ile
                                                    1870
                                1865
            1860
Phe Pro Phe Asp Tyr Leu Pro Ala Glu Gln Val Cys Thr Ile Ala Lys
                                                1885
                            1880
        1875
Lys Asp Ala Phe Trp Arg Leu Asp Lys Thr Glu Ser Lys Ile Pro Ala
                                            1900
                        1895
    1890
Arg Val Val Phe Gln Ile Trp Asp Asn Asp Lys Phe Ser Phe Asp Asp
                                                            1920
                                        1915
                    1910
1905
Phe Leu Gly Ser Leu Gln Leu Asp Leu Asn Arg Met Pro Lys Pro Ala
                                                        1935
                                    1930
                1925
Lys Thr Ala Lys Lys Cys Ser Leu Asp Gln Leu Asp Asp Ala Phe His
                                1945
                                                    1950
            1940
Pro Glu Trp Phe Val Ser Leu Phe Glu Gln Lys Thr Val Lys Gly Trp
                                                 1965
                            1960
        1955
Trp Pro Cys Val Ala Glu Glu Gly Glu Lys Lys Ile Leu Ala Gly Lys
                                             1980
                        1975
    1970
Leu Glu Met Thr Leu Glu Ile Val Ala Glu Ser Glu His Glu Glu Arg
                                                             2000
                                         1995
                    1990
1985
Pro Ala Gly Gln Gly Arg Asp Glu Pro Asn Met Asn Pro Lys Leu Glu
                                                         2015
                                    2010
                2005
Asp Pro Arg Arg Pro Asp Thr Ser Phe Leu Trp Phe Thr Ser Pro Tyr
                                                     2030
                                2025
            2020
Lys Thr Met Lys Phe Ile Leu Trp Arg Arg Phe Arg Trp Ala Ile Ile
                                                 2045
                            2040
        2035
Leu Phe Ile Ile Leu Phe Ile Leu Leu Phe Leu Ala Ile Phe Ile
                                             2060
                        2055
    2050
Tyr Ala Phe Pro Asn Tyr Ala Ala Met Lys Leu Val Lys Pro Phe Ser
                                         2075
                     2070
 2065
       <210> 3
       <211> 5915
       <212> DNA
       <213> Homo sapiens
```

<400> 3 tcgaccgccc agccaggtgc aaaatgccgt gtcattggga gactccgcag ccggagcatt agattacage tegacggage tegggaaggg eggeggggt ggaagatgag eagaageece tgttetegga acgeeggetg acaagegggg tgaagegeagg eggggegggg acceageeta geecactgga geageegggg gtggeeegtt eccetttaag ageaactget etaageeagg agccagagat tegageegge etegeceage cageeetete cagegagggg acceacaage ggcgcetegg ccetecegae etticegage cetetttgeg ccetgggege acggggecet acacgegeca ageatgetga gggtetteat ectetatgee gagaacgtee acacaceega caccgacatc agcgatgcct actgctccgc ggtgtttgca ggggtgaaga agagaaccaa agtcatcaag aacagcgtga accetgtatg gaatgaggga tttgaatggg accetaaggg catcecetg gaccagget etgagettea tgtggtggte aaagaccatg agacgatggg gaggaacagg ttcctggggg aagccaaggt cccactccga gaggtcctcg ccacccctag tetgteegee agetteaatg ecceetget ggacaccaag aageageeca caggggeete getggteetg caggtgteet acacaceget geetggaget gtgeecetgt teeegeeeee tacteetetg gageeeteee egactetgee tgacetggat gtagtggeag acacaggagg agaggaagac acagaggace agggacteae tggagatgag geggageeat teetggatea aageggagge cegggggete ceaccacece aaggaaacta cetteaegte eteegeecea ctacccggg atcaaaagaa agcgaagtgc gcctacatct agaaagctgc tgtcagacaa accgcaggat ttccagatca gggtccaggt gatcgagggg cgccagctgc cggggggtgaa catcaageet gtggteaagg ttaccgetge agggeagace aageggaege ggatecacaa gggaaacage ccactettea atgagactet tttetteaae ttgtttgaet eteetgggga getgtttgat gageceatet ttateaeggt ggtagaetet egitetetea ggacagaige tetecteggg gagtteegga tggaegtggg caccatttae agagageece ggeaegeeta teteaggaag tggetgetge teteagaece tgatgaette tetgetgggg ecagaggeta cctgaaaaca agcctttgtg tgctggggcc tggggacgaa gcgcctctgg agagaaaaga 

cccctctgaa gacaaggagg acattgaaag caacctgctc cggcccacag gcgtagccct gcgaggagcc cacttctgcc tgaaggtctt ccgggccgag gacttgccgc agatggacga tgccgtgatg gacaacgtga aacagatctt tggcttcgag agtaacaaga agaacttggt ggaccccttt gtggaggtca gctttgcggg gaaaatgctg tgcagcaaga tcttggagaa gacggccaac cctcagtgga accagaacat cacactgcct gccatgtttc cctccatgtg cgaaaaatg aggattegta teatagactg ggacegeetg acteacaatg acategtgge taccacetae etgagtatgt egaaaatete tgeeeetgga ggagaaatag aagaggagee tgcaggtgct gtcaagcctt cgaaagcctc agacttggat gactacctgg gcttcctccc cacttttggg ccctgctaca tcaacctcta tggcagtccc agaggttca caggcttccc agacccctac acaggctca acacaggcaa gggggaaggt gtggcttatc gtggccggct tctgctctcc ctggagacca agctggtgga gcacagtgaa cagaaggtgg aggaccttcc tgcggatgac atcctccggg tggagaagta ccttaggagg cgcaagtact ccctgtttgc ggccttctac tcagccacca tgctgcagga tgtggatgat gccatccagt ttgaggtcag catcgggaac tacgggaaca agttcgacat gacctgcctg ccgctggcct ccaccactca gtacagccgt gcagtctttg acgggtgcca ctactactac ctaccctggg gtaacgtgaa acctgtggtg gtgctgtcat cctactggga ggacatcagc catagaatcg agactcagaa ccagctgctt gggattgctg accggctgga agctggcctg gagcaggtcc acctggccct gaaggcgcag tgctccacgg aggacgtgga ctcgctggtg gctcagctga cggatgagct catcgcaggc tgcagccagc ctctgggtga catccatgag acaccctctg ccacccacct ggaccagtac ctgtaccage tgcgcaccca tcacctgage caaatcactg aggetgeeet ggcctgaag ctcggccaca gtgagctccc tgcagctctg gagcaggcgg aggactggct cctgcgtctg cgtgccctgg cagaggagcc ccagaacagc ctgccggaca tcgtcatctg gatgctgcag ggagacaagc gtgtggcata ccagcgggtg cccgccacc aagtcctctt ctcccggcgg ggtgccaact actgtggcaa gaattgtggg aagctacaga caatctttct gaaatatccg atggagaagg tgcctggcgc ccggatgcca gtgcagatac gggtcaagct gtggtttggg ctctctgtgg atgagaagga gttcaaccag tttgctgagg ggaagctgtc tgtctttgct gaaacctatg agaacgagac taagttggcc cttgttggga actggggcac aacgggcctc acctacccca agttttctga cgtcacgggc aagatcaagc tacccaagga cagcttccgc cctcggccg gctggacctg ggctggagat tggttcgtgt gtccggagaa gactctgctc catgacatgg acgccggtca cctgagcttc gtggaagagg tgtttgagaa ccagacccgg cttcccggag gccagtggat ctacatgagt gacaactaca ccgatgtgaa cggggagaag gtgcttccca aggatgacat tgagtgccca ctgggctgga agtgggaaga tgaggaatgg tccacagacc tcaaccgggc tgtcgatgag caaggctggg agtatagcat caccatccc ccggagcgga agccgaagca ctgggtccct gctgagaaga tgtactacac acaccgacgg cggcgctggg tgcgcctgcg caggagggat ctcagccaaa tggaagcact gaaaaggcac aggcaggcgg aggcggaggg cgagggctgg gagtacgcct ctctttttgg ctggaagttc cacctcgagt accgcaagac agatgccttc cgccgccgcc gctggcgccg tegeatggag ceaetggaga agaeggggee tgeagetgtg titgeeettg agggggeeet gggeggegtg atggatgaca agagtgaaga tteeatgtee gteteeaeet tgagettegg tgtgaacaga cccacgattt cctgcatatt cgactatggg aaccgctacc atctacgctg ctacatgtac caggeceggg acctggetge gatggacaag gactetttt etgateceta tgccatcgtc tccttcctgc accagagcca gaagacggtg gtggtgaaga acacccttaa cccacctgg gaccagacge teatetteta egagategag atetttggeg agecggecae agttgetgag caacegecea geattgtggt ggagetgtae gaccatgaea ettatggtge agacgagttt atgggteget geatetgtea accgagtetg gaacggatge caeggetgge ctggttccca ctgacgaggg gcagccagcc gtcgggggag ctgctggcct cttttgagct catccagaga gagaagccgg ccatccacca tattcctggt tttgaggtgc aggagacatc aaggateetg gatgagtetg aggacacaga cetgecetae ceaceaece agagggagge caacatetae atggtteete agaacateaa gecagegete cagegtaeeg ceategagat cctggcatgg ggcctgcgga acatgaagag ttaccagctg gccaacatct cctccccag cctcgtggta gagtgtgggg gccagacggt gcagtcctgt gtcatcagga acctccggaa gaaccccaac tttgacatct gcaccctctt catggaagtg atgctgccca gggaggagct ctactgccc cccatcaccg tcaaggtcat cgataaccgc cagtttggcc gccggcctgt ggtgggccag tgtaccatcc gctccctgga gagcttcctg tgtgacccct actcggcgga gagtccatcc ccacagggtg gcccagacga tgtgagccta ctcagtcctg gggaagacgt gctcatcgac attgatgaca aggagccct catccccatc caggaggaag agttcatcga ttggtggage aaattetttg cetecatagg ggagagggaa aagtgegget cetacetgga gaaggatttt gacacetga aggtetatga cacacagetg gagaatgtgg aggeetttga gggcctgtct gacttttgta acaccttcaa gctgtaccgg ggcaagacgc aggaggagac agaagatcca tctgtgattg gtgaatttaa gggcctcttc aaaatttatc ccctcccaga agacccagcc atccccatgc ccccaagaca gttccaccag ctggccgccc agggacccca ggagtgcttg gtccgtatct acattgtccg agcatttggc ctgcagccca aggaccccaa tggaaagtgt gatccttaca tcaagatctc catagggaag aaatcagtga gtgaccagga taactacatc ccctgcacgc tggagcccgt atttggaaag atgttcgagc tgacctgcac totgoototg gagaaggaco taaagatoao tototatgao tatgacotoo totocaagga cgaaaagatc ggtgagacgg tcgtcgacct ggagaacagg ctgctgtcca agtttggggc tcgctgtgga ctcccacaga cctactgtgt ctctggaccg aaccagtggc gggaccagct 

ccggacagac caggatccca gcagcagggc agacatcgag gcctggacct tatctggaat	cagetectee acetettetg e egtgtaatgt tteaggataa a aacecacace tgggeecagt g etggteeegg ageacgtgga g caggggaage tgeagatgtg g ecetteaaca teacecacg g aceagagatg tgateetgga t gtgaaaggtt ggatgattgg e	ggagtattcc ggaggagcgt gtcacggccc ggtcgaccta gagagccaga cgacctgagc	attgaagaga ctggctctgc ctctacagcc tttccgaagg aggtttttcc	atgtgcttca ccctgcagcc ccctggggcg tgcgttgtat	5520 5580 5640 5700 5760 5820 5880 5915
<210 <21	> 4 > 20				
	> DNA > Homo sapiens				
<400 tgggacctca	> 4 aagggcatcc				20
<210	)> 5	•			
<21	.> 20				
	<pre>Post</pre>				
<40	)> 5				
	aggatgtgga				20
	)> 6				
	L> 20 2> DNA				
	3> Homo sapiens				
<40	)> 6				20
gggaggtga	a gcaacttcaa				20
	)> 7				
·—·—	1> 20				
	2> DNA 3> Homo sapiens				
<40	0> 7				
	agaagatgag				20
<21	0> 8				
<21	1> 20				
	2> DNA				
<21	3> Homo sapiens				
<40	0> 8				20
cagggccga	g atgagcccaa				20
	0> 9				
	1> 20				
	2> DNA 3> Homo sapiens				
<40	0> 9				
	g tcctggatga				20
	0> 10				
	1> 20				
	2> DNA 3> Homo sapiens				
	0> 10				20
ctgtggcgg	t gtttccggtg				20
	0> 11				
<21	1> 20				

```
<212> DNA
        <213> Homo sapiens
        <400> 11
                                                                                          20
 acagacgtgc gttatcgttc
        <210> 12
        <211> 20
        <212> DNA
        <213> Homo sapiens
        <400> 12
                                                                                          20
 aagactgagc aaaatcccag
        <210> 13
        <211> 6912
         <212> DNA
         <213> Homo sapiens
         <400> 13
tcgaccgccc agccaggtgc aaaatgccgt gtcattggga gactccgcag ccggagcatt
                                                                                         60
                                                                                        120
agattacago togacggago togggaaggg oggogggggt ggaagatgag cagaagcoco
                                                                                        180
gcccactgga gcagccggg gtggcccgtt cccctttaag agcaactgct ctaagccagg agccagagat tcgagccggc ctcgcccagc cagccctctc cagcgagggg acccacaagc
                                                                                        240
                                                                                        300
ggegeetegg ceeteegae etticegage cetettgeg ceetgggege acggggeet
                                                                                        360
acacgogoca agcatgotga gggtottcat cototatgoc gagaacgtoc acacacocga
                                                                                        420
caccgacatc agcgatgcct actgctccgc ggtgtttgca ggggtgaaga agagaaccaa
                                                                                        480
agtcatcaag aacagcgtga accetgtatg gaatgaggga titgaatggg acctcaaagg gcatcccct ggaccaggge tetgagette atgtggtggt caaagaccat gagacgatgg
                                                                                        540
                                                                                        600
ggaggaacag gttcctgggg gaagccaagg tcccactccg agaggtcctc gccacccta
                                                                                        660
gtetgteege cagetteaat geeceeetge tggacaceaa gaageageee acaggggeet
                                                                                        720
egetggteet geaggtgtee tacacacege tgeetggage tgtgeecetg tteeegeece
                                                                                        780
                                                                                        840
ctactectet ggageeetee eegactetge etgacetgga tgtagtggea gacacaggag
gagaggaaga cacagaggac cagggactca ctggagatga ggcggagcca ttcctggatc aaagcggagg cccgggggct cccaccacc caaggaaact acettcacgt cetcegccc
                                                                                        900
                                                                                        960
actacccgg gatcaaaaga aagcgaagtg cgcctacatc tagaaagctg ctgtcagaca aaccgcagga tttccagatc agggtccagg tgatcgaggg gcgccagctg ccgggggtga
                                                                                       1020
                                                                                       1080
acatcaagcc tgtggtcaag gttaccgctg cagggcagac caagcggacg cggatccaca agggaaacag cccactcttc aatgagactc ttttcttcaa cttgtttgac tctcctgggg agctgtttga tgagcccatc tttatcacgg tggtagactc tcgttctctc aggacagatg
                                                                                       1140
                                                                                       1200
                                                                                       1260
ctetectegg ggagtteegg atggaegtgg gcaccattta cagagagece eggeacgeet
                                                                                       1320
atctcaggaa gtggctgctg ctctcagacc ctgatgactt ctctgctggg gccagaggct
                                                                                       1380
acctgaaaac aagcctttgt gtgctggggc ctggggacga agcgcctctg gagagaaaag
                                                                                       1440
accetetga agacaaggag gacattgaaa geaacetget ceggeecaca ggegtageec tgegaggage ceaettetge etgaaggtet teegggeega ggaettgeeg cagatggaeg
                                                                                       1500
                                                                                       1560
atgccgtgat ggacaacgtg aaacagatct ttggcttcga gagtaacaag aagaacttgg
tggaccctt tgtggaggtc agctttgcgg ggaaaatgct gtgcagcaag atcttggaga
                                                                                       1620
                                                                                       1680
agacggccaa ccctcagtgg aaccagaaca tcacactgcc tgccatgttt ccctccatgt
                                                                                       1740
                                                                                       1800
gcgaaaaaat gaggattegt atcatagact gggacegeet gaetcacaat gacategtgg
ctaccaccta cctgagtatg tcgaaaatct ctgcccctgg aggagaaata gaagaggagc ctgcaggtgc tgtcaagcct tcgaaagcct cagacttgga tgactacctg ggcttcctcc
                                                                                       1860
                                                                                       1920
 ccacttttgg gccctgctac atcaacctct atggcagtcc cagagagttc acaggcttcc
                                                                                       1980
 cagaccccta cacagagete aacacaggea agggggaagg tgtggettat cgtggccgge
                                                                                       2040
                                                                                       2100
 ttctgctctc cctggagacc aagctggtgg agcacagtga acagaaggtg gaggaccttc
 ctgcggatga catcctccgg gtggagaagt accttaggag gcgcaagtac tccctgtttg cggccttcta ctcagccacc atgctgtagg atgtggatga tgccatccag tttgaggtca
                                                                                       2160
                                                                                       2220
 gcatcgggaa ctacgggaac aagttcgaca tgacctgcct gccgctggcc tccaccactc
                                                                                       2280
                                                                                       2340
 agtacagccg tgcagtcttt gacgggtgcc actactacta cctaccctgg ggtaacgtga
                                                                                       2400
 aacctgtggt ggtgctgtca tcctactggg aggacatcag ccatagaatc gagactcaga
 accagetget tgggattget gaccggetgg aagetggeet ggagcaggte cacetggeee
                                                                                       2460
 tgaaggegea gigeteeacg gaggaegtgg actegeiggt ggeteagetg acggaigage
                                                                                       2520
 teategeagg etgeageeag ectetgggtg acatecatga gacacectet gecacecace
                                                                                        2580
 tggaccagta cetgtaccag etgegeacce ateacetgag ceaaateact gaggetgece
                                                                                       2640
 tggccctgaa gctcggccac agtgagctcc ctgcagctct ggagcaggcg gaggactggc
                                                                                       2700
 tcctgcgtct gcgtgccctg gcagaggagc cccagaacag cctgccggac atcgtcatct ggatgctgca gggagacaag cgtgtggcat accagcgggt gcccgcccac caagtcctct
                                                                                        2760
                                                                                        2820
```

tctcccggcg gggtgccaac tactgtggca agaattgtgg gaagctacag acaatctttc tgaaatatcc gatggagaag gtgcctggcg cccggatgcc agtgcagata cgggtcaagc tgtggtttgg gctctctgtg gatgagaagg agttcaacca gtttgctgag gggaagctgt ctgtctttgc tgaaacctat gagaacgaga ctaagttggc ccttgttggg aactggggca caacgggcet cacctaccc aagttttetg acgteacggg caagateaag ctacccaagg acagettecg ecceteggee ggetggacet gggetggaga ttggttegtg tgteeggaga agaetetget ceatgacatg gaegeeggte acetgagett egtggaagag gtgtttgaga accagacccg gcttcccgga ggccagtgga tctacatgag tgacaactac accgatgtga acggggagaa ggtgcttccc aaggatgaca ttgagtgccc actgggctgg aagtgggaag atgaggaatg gtccacagac ctcaaccggg ctgtcgatga gcaaggctgg gagtatagca tcaccatccc cccggagcgg aagccgaagc actgggtccc tgctgagaag atgtactaca cacaccgacg geggegetgg gtgegeetge geaggaggga teteageeaa atggaageae tgaaaaggca caggcaggcg gaggcggagg gcgagggctg ggagtacgcc tctcttttg gctggaagtt ccacctcgag taccgcaaga cagatgcctt ccgccgccgc cgctggcgcc gtcgcatgga gccactggag aagacggggc ctgcagctgt gtttgccctt gagggggccc tggggcgcgt gatggatgac aagagtgaag attccatgtc cgtctccacc ttgagcttcg gtgtgaacag acccacgatt tectgcatat tegactatgg gaacegetac catetacget gctacatgta ccaggcccgg gacctggctg cgatggacaa ggactctttt tctgatccct atgccatcgt ctccttcctg caccagagcc agaagacggt ggtggtgaag aacaccctta accecacetg ggaceagacg etcatettet acgagatega gatetttgge gageeggeea cagttgctga gcaaccgccc agcattgtgg tggagctgta cgaccatgac acttatggtg cagacgagtt tatgggtcgc tgcatctgtc aaccgagtct ggaacggatg ccacggctgg cctggttccc actgacgagg ggcagccagc cgtcgggga gctgctggcc tcttttgagc caaggatcct ggatgagtct gaggacacag acctgccta tttgaggtg caggagagagcccaacatcta catggttcct cagaacatca agccagcgct ccagcgtacc gccatcgaga tectggeatg gggeetgegg aacatgaaga gttaccaget ggeeaacate tecteecea gcetegtggt agagtgtggg ggccagacgg tgcagtcetg tgtcatcagg aacetecgga agaaceccaa etttgacate tgcacectet tcatggaagt gatgetgeec agggaggage tctactgccc ccccatcacc gtcaaggtca tcgataaccg ccagtttggc cgccggcctg tggtgggcca gtgtaccatc cgctccctgg agagettect gtgtgacccc tactcggcgg agagtecate eccaeagget geccagaeg atgtgageet acteagteet ggggaagaeg tgctcatcga cattgatgac aaggagccc tcatcccat ccaggaggaa gagttcatcg attggtggag caaattcttt gcctccatag gggagaggga aaagtgcggc tcctacctgg agagggcttt tgacaccttg aaggtctatg acacacagct gggaaatgtg gaggcctttg agggcctgtc tgacttttgt aacaccttca aggtgtaaccg gggcaagacg caggaggaga cagaagatcc atctgtgatt ggtgaattta agggcctctt caaaatttat cccctcccag aagacccage catecccatg cccccaagac agttecacca getggeegee cagggacccc aggagtgett ggteegtate tacattgtee gageatttgg cetgeageee aaggaceeea atggaaagtg tgateettae ateaagatet ecatagggaa gaaateagtg agtgaceagg ataactacat cccctgcacg ctggagcccg tatttggaaa gatgttcgag ctgacctgca ctctgcctct ggagaaggac ctaaagatca ctctctatga ctatgacctc ctctccaagg acgaaaagat cggtgagacg gtcgtcgacc tggagaacag gctgctgtcc aagtttgggg ctcgctgtgg actccacag acctactgtg tctctggacc gaaccagtgg cgggaccagc tccgccctc ccacctctct cacctcttct gccagcagca tagagtcaag gcacctgtgt accggacaga ccgtgtaatg tttcaggata aagaatattc cattgaagag atagaggctg gcaggatccc aaacccacac ctgggcccag tggaggagcg tctggctctg catgtgcttc agcagcaggg cctggtcccg gagcacgtgg agtcacggcc cctctacagc cccctgcagc cagacatcga gcaggggaag ctgcagatgt gggtcgacct atttccgaag gccctggggc ggcctggacc tcccttcaac atcaccccac ggagagccag aaggtttttc ctgcgttgta ttatctggaa taccagagat gtgatcctgg atgacctgag cctcacgggg gagaagatga gcgacattta tgtgaaaggt tggatgattg gctttgaaga acacaagcaa aagacagacg tgcattatcg ttccctggga ggtgaaggca acttcaactg gaggttcatt ttccccttcg actacctgcc agctgagcaa gtctgtacca ttgccaagaa ggatgccttc tggaggctgg acaagactga gagcaaaatc ccagcacgag tggtgttcca gatctgggac aatgacaagt tctcctttga tgattttctg ggctccctgc agctcgatct caaccgcatg cccaagacagc ccaagacagc ccaagactgc tccttggacc agctggatga tgctttccac ccaagactgt ttgtgtccct ttttgagcag aaaacagtga agggctggtg gccctgtgta gcagaagagggtgagaagaa aatactggcg ggcaagctgg aaatgacctt ggagattgta gcagagagtg agcatgagga geggeetget ggeeagggee gggatgagee caacatgaac cetaagettg aggacecaag gegeeegac aceteettee tgtggtttac etceecatac aagaceatga agtteateet gtggeggegt tteeggtggg ceatcateet etteateate etetteatee gattgtcctg ccagggtggg cagacagaca gatggaccgg cccacactcc cagagttgct aacatggagc tctgagatca cccacttcc atcatttcct tctcccccaa cccaacgctt 

6900

#### 18/68

6912 aaaaaaaaa aa <210> 14 <211> 6911 <212> DNA <213> Homo sapiens <400> 14 60 tcgaccgccc agccaggtgc aaaatgccgt gtcattggga gactccgcag ccggagcatt 120 agattacago togacggago togggaaggg oggogggggt ggaagatgag cagaagcooc 180 240 300 ggcgcctcgg ccctcccgac ctttccgagc cctctttgcg ccctgggcgc acggggccct 360 420 acacqcqcca agcatgctga gggtcttcat cctctatgcc gagaacgtcc acacacccga caccgacatc agcgatgct actgctccgc ggtgtttgca ggggtgaaga agagaaccaa agtcatcaag aacagcgtga accctgtatg gaatgaggga tttgaatggg acctcaaggg catcccctg gaccagggct ctgagcttca tgtggtggtc aaagaccatg agacgatggg 480 540 600 gaggaacagg ttcctggggg aagccaaggt cccactccga gaggtcctcg ccaccctag 660 720 tetatecace agetteaatg ecceetget ggacaceaag aageageeca caggggeete gctggtcctg caggtgtcct acacaccgct gcctggagct gtgcccctgt tcccgccccc tactcctctg gagccctccc cgactctgcc tgacctggat gtagtggcag acacaggagg agaggaagac acagaggacc tggagatgag gcggagccat tcctggatca 780 840 900 aagcggagge cegggggete ceaceacec aaggaaacta cetteacgte etecgecea 960 ctaccceggg atcaaaagaa agcgaagtgc gcctacatct agaaagctgc tgtcagacaa 1020 1080 accqcaggat ttccagatca gggtccaggt gatcgagggg cgccagctgc cgggggtgaa catcaageet gtggteaagg ttacegetge agggeagace aageggaege ggateeacaa gggaaacage ceactettea atgagaetet tttetteaac ttgtttgaet eteetgggga 1140 1200 gctgtttgat gagcccatct ttatcacggt ggtagactct cgttctctca ggacagatgc 1260 tetecteggg gagtteegga tggaegtggg caccatttae agagageece ggeaegeeta 1320 1380 teteaggaag tggetgetge teteagacce tgatgaette tetgetgggg ccagaggeta cctgaaaaca agcctttgtg tgctggggcc tggggacgaa gcgcctctgg agagaaaaga cccctctgaa gacaaggagg acattgaaag caacctgctc cggcccacag gcgtagccct gcgaggagcc cacttctgcc tgaaggtctt ccgggccgag gacttgccgc agatggacga 1440 1500 1560 tgccqtqatg gacaacgtga aacagatctt tggcttcgag agtaacaaga agaacttggt 1620 ggaccccttt gtggaggtca gctttgcggg gaaaatgctg tgcagcaaga tcttggagaa gacggccaac cctcagtgga accagaacat cacactgcct gccatgtttc cctccatgtg 1680 1740 cgaaaaaatg aggattegta teatagaetg ggacegeetg acteacaatg acategtgge taccacetae etgagtatgt egaaaatete tgeeeetgga ggagaaatag aagaggagee 1800 1860 tgcaggtgct gtcaagcctt cgaaagcctc agacttggat gactacctgg gcttcctccc 1920 cacttitggg ccctgctaca tcaacctcta tggcagtccc agagagttca caggcttccc 1980 agaccectae acagagetea acacaggeaa gggggaaggt gtggettate gtggeegget tetgetetee etggagaeca agetggtgga geacagtgaa cagaaggtgg aggacettee 2040 2100 tgcggatgac atcctccggg tggagaagta ccttaggagg cgcaagtact ccctgtttgc 2160 ggccttctac tcagccacca tgctgtagga tgtggatgat gccatccagt ttgaggtcag 2220 2280 catcoggaac tacoggaaca acttogacat gacctoccto cogetogeet ccaccactea gtacagccgt gcagtctttg acgggtgcca ctactactac ctaccctggg gtaacgtgaa 2340 acctgtggtg gtgctgtcat cctactggga ggacatcagc catagaatcg agactcagaa 2400 ccagctgctt gggattgctg accggctgga agctggcctg gagcaggtcc acctggccct 2460 gaaggegeag tgeteeacgg aggaegtgga etegetggtg geteagetga eggatgaget 2520 categoagge tgcagecage etetgggtga catecatgag acaecetetg ceacceacet 2580 ggaccagtac ctgtaccagc tgcgcaccca tcacctgagc caaatcactg aggctgccct 2640 ggccctgaag ctcggccaca gtgagctccc tgcagctctg gagcaggcgg aggactggct cctgcgtctg cgtgccctgg cagaggagcc ccagaacagc ctgccggaca tcgtcatctg 2700 2760 gatgetgeag ggagacaage gtgtggeata ccagegggtg ccegeceace aagteetett 2820 ctcccggcgg ggtgccaact actgtggcaa gaattgtggg aagctacaga caatctttct 2880 gaaatateeg atggagaagg tgeetggege eeggatgeea gtgeagatae gggteaaget 2940 gtggtttggg ctctctgtgg atgagaagga gttcaaccag tttgctgagg ggaagctgtc tgtctttgct gaaacctatg agaacgagac taagttggcc cttgttggga actggggcac 3000 3060 aacgggcctc acctacccca agttttctga cgtcacgggc aagatcaagc tacccaagga 3120 cagettecge cecteggeeg getggaeetg ggetggagat tggttegtgt gteeggagaa 3180 gactctgctc catgacatgg acgccggtca cctgagcttc gtggaagagg tgtttgagaa ccagacccgg cttcccgag gccagtggat ctacatgagt gacaactaca ccgatgtgaa cggggagaag gtgcttccca aggatgacat tgagtgccca ctgggctgga agtgggaaga 3240 3300 3360 3420 tgaggaatgg tccacagacc tcaaccgggc tgtcgatgag caaggctggg agtatagcat caccatece ceggagegga ageegaagea etgggteeet getgagaaga tgtactacae 3480

3540

#### 19/68

```
acaccgacgg cggcgctggg tgcgcctgcg caggagggat ctcagccaaa tggaagcact
gaaaaggcac aggcaggcgg aggcggaggg cgagggctgg gagtacgcct ctctttttggctggaagttc cacctcgagt accgcaagac agatgccttc cgccgccgcc gctggcgccg
                                                                                                                 3600
                                                                                                                 3660
tegeatggag ceaetggaga agaeggggee tgeagetgtg tttgeeettg agggggeeet
                                                                                                                 3720
gggcggcgtg atggatgaca agagtgaaga ttccatgtcc gtctccacct tgagcttcgg
                                                                                                                 3780
tgtgaacaga cccacgattt cctgcatatt cgactatggg aaccgctacc atctacgctg
                                                                                                                 3840
ctacatgtac caggeocggg acctggetge gatggacaag gactetttt etgateeta tgecategte teetteetge accagageea gaagaeggtg gtggtgaaga acaceettaa
                                                                                                                 3900
                                                                                                                 3960
cccacctgg gaccagacgc tcatcttcta cgagatcgag atctttggcg agccggccac
                                                                                                                 4020
agttgctgag caaccgccca gcattgtggt ggagctgtac gaccatgaca cttatggtgc
                                                                                                                 4080
agacgagttt atgggtcgct gcatctgtca accgagtctg gaacggatgc cacggctggc ctggttcca ctgacgaggg gcagccagcc gtcggggag ctgctggcct cttttgagct catccagaga gagaagccgg ccatccacca tattcctggt tttgaggtgc aggagacatc
                                                                                                                 4140
                                                                                                                 4200
                                                                                                                 4260
aaggatcetg gatgagtetg aggacacaga cetgeeetac ceaceacee agagggagge caacatetac atggtteete agaacateaa geeagegete cagegtaceg ceategagat cetegtggta gagtgtgggg geeagaeggt geagteetgt gteateagga aceteeggaa gaaceceaac tttgacatet geacectett catggaagtg atgetgees gggaggaget ctaetgees ceateaceg teagaggagget catggaagtg atgetgees gggaggaget
                                                                                                                 4320
                                                                                                                 4380
                                                                                                                 4440
                                                                                                                 4500
                                                                                                                 4560
ctactgccc cccatcaccg tcaaggtcat cgataaccgc cagtttggcc gccggcctgt ggtgggccag tgtaccatcc gctccctgga gagcttcctg tgtgacccct actcggcgga
                                                                                                                  4620
                                                                                                                  4680
gagtecatec ceacagggtg geceagaega tgtgageeta eteagteetg gggaagaegt geteategae attgatgaea aggageeeet catececate caggaggaag agtteatega
                                                                                                                  4740
                                                                                                                  4800
ttggtggagc aaattetttg cetecatagg ggagagggaa aagtgegget cetacetgga
gaaggatttt gacaceetga aggtetatga cacacagetg gagaatgtgg aggeetttga
                                                                                                                  4860
                                                                                                                  4920
gggcctgtct gacttttgta acaccttcaa gctgtaccgg ggcaagacgc aggaggagac agaagatcca tctgtgattg gtgaatttaa gggcctcttc aaaatttatc ccctcccaga
                                                                                                                  4980
                                                                                                                  5040
                                                                                                                  5100
agaccagec atcccatge eccaagaca gttecaccag etggeegeec agggacceca
ggagtgettg gteegtatet acattgteeg ageatttgge etgeageeca aggaceecaa tggaaagtgt gateettaca teaagatete catagggaag aaateagtga gtgaceagga
                                                                                                                  5160
                                                                                                                  5220
 taactacatc ccctgcacgc tggagcccgt atttggaaag atgttcgagc tgacctgcac
                                                                                                                  5280
tetgeetetg gagaaggace taaagateae tetetatgae tatgacetee tetecaagga egagaaggae egegageet ggagaacagg etgetgteea agtttggge eegeetete eageteetee eageteetee acetettetg eeageageat agagteaagg eacetgtgta
                                                                                                                  5340
                                                                                                                  5400
                                                                                                                  5460
                                                                                                                  5520
 ccggacagac cgtgtaatgt ttcaggataa agaatattcc attgaagaga tagaggctgg
                                                                                                                  5580
 caggatecca aacccacace tgggcccagt ggaggagegt etggetetge atgtgettea
                                                                                                                  5640
 gcagcaggge ctggtcccgg agcacgtgga gtcacggccc ctctacagcc ccctgcagcc agacatcgag caggggaagc tgcagatgtg ggtcgaccta tttccgaagg ccctggggcg
                                                                                                                  5700
                                                                                                                  5760
 gcctggacet cccttcaaca teaccccacg gagagecaga aggtttttcc tgcgttgtat tatctggaat accagagatg tgatcctgga tgacctgage ctcacggggg agaagatgag
                                                                                                                  5820
                                                                                                                  5880
 cgacatttat gtgaaaggtt ggatgattgg ctttgaagaa cacaagcaaa agacagacgt
                                                                                                                  5940
 gcattatcgt tccctgggag gtgaaggcaa cttcaactgg aggttcattt tccccttcga
                                                                                                                  6000
 ctacctgcca gctgagcaag tctgtaccat tgccaagaag gatgccttct ggaggctgga caagactgag agcaaaatcc cagcacgagt ggtgttccag atctgggaca atgacaagtt
                                                                                                                  6060
                                                                                                                  6120
 ctcctttgat gattttctgg gctccctgca gctcgatctc aaccgcatgc ccaagccagc caagacagcc aagaagtgct ccttggacca gctggatgat gctttccacc cagaatggtt
                                                                                                                  6180
                                                                                                                   6240
                                                                                                                   6300
 tqtqtccctt tttgagcaga aaacagtgaa gggctggtgg ccctgtgtag cagaagaggg
 tgagaagaaa atactggcgg gcaagctgga aatgaccttg gagattgtag cagaaggtga gcatgaggag cgcctgctg gccagggccg ggatgagccc aacatgaacc ctaagcttga ggacccaagg cgcccgaca cctccttcct gtggtttacc tccccataca agaccatgaa gttcatcctt tggcggcgtt tccggtgggc catcatcctc ttcatcatcct
                                                                                                                   6360
                                                                                                                   6420
                                                                                                                   6480
                                                                                                                   6540
  getgetgtte etggecatet teatetacge ettecegaae tatgetgeca tgaagetggt
                                                                                                                   6600
  gaageeette agetgaggae teteetgeee tgtagaaggg geegtggggt eeeeteeage atgggaetgg cetgeeteet eegeeeaget eggegagete etceagaeet eetaggeetg
                                                                                                                   6660
                                                                                                                   6720
  attgtectge cagggtggge agacagacag atggacegge ceacactece agagttgeta
                                                                                                                   6780
  acatggaget etgagateae eccactteca teattteett etecceaae ecaaegettt
                                                                                                                   6840
  tttggatcag ctcagacata tttcagtata aaacagttgg aaccacaaaa aaaaaaaaa
                                                                                                                   6900
                                                                                                                   6911
  aaaaaaaaa a
```

```
<210> 15
<211> 6910
```

<212> DNA

<213> Homo sapiens

<400> 15

tcgaccgccc agccaggtgc aaaatgccgt gtcattggga gactccgcag ccggagcatt 60 agattacage tegacggage tegggaaggg eggegggggt ggaagatgag cagaageece 120

tgttctcgga	acgccggctg	acaagcgggg	tgagcgcagg	caaaacaaaa	acccageeta	180
gcccactgga	gcagccgggg	ataacccatt	cccctttaag	aggaagtggt	ctaaccacc	240
anccananat	tcgagccggc	ctccccacc	caccatata	agouactge:	ccaagccagg	
agecagagae		cttgcctage	t-t-t-	cagegagggg	acceacaage	300
ggegeetegg	ccctcccgac	cttteegage	cctctttgcg	ccctgggcgc	acggggccct	360
acacgcgcca	agcatgctga	gggtcttcat	cctctatgcc	gagaacgtcc	acacacccga	420
caccgacatc	agcgatgcct	actgctccgc	ggtgtttgca	ggggtgaaga	agagaaccaa	480
agtcatcaag	aacagcgtga	accctgtatg	gaatgaggga	tttgaatggg	acctcaaggg	540
catececeta	gaccagggct	ctanacttan	tataataata	22222222	accedaggg	
anaan nanaa	ttaataaaa	anagettea	egeggeggee	adagactatg	agacgatggg	600
gaggaacagg	ttcctggggg	aagecaaggt	cccactccga	gaggteeteg	ccacccctag	660
tctgtccgcc	agcttcaatg	ccccctgct	ggacaccaag	aagcagccca	caggggcctc	720
gctggtcctg	caggtgtcct	acacaccgct	gcctggagct	gtgcccctgt	teccacacac	780
tactcctctq	gagccctccc	cgactctgcc	tgacctggat	gtagtgggag	acacaggagg	840
anannaanan	acagaggacc	agggagtgag	tocacatoac	gongoggong	teacaggagg	
agaggaagac	acagaggace	agggacteat	cggagacgag	geggageeat	teetggatea	900
aageggagge	ccgggggctc	ccaccacccc	aaggaaacta	ccttcacgtc	ctccgcccca	960
ctaccccggg	atcaaaagaa	agcgaagtgc	gcctacatct	agaaagctgc	tgtcagacaa	1020
accgcaggat	ttccagatca	gggtccaggt	gatcgagggg	caccaactac	cagagataaa	1080
catcaagect	gtggtcaagg	ttaccactac	aggggagacc	aagggaggg	ccatccacaa	1140
addaaacaac	ccactcttca	atragactet	tttcttcaac	ttatttasat	stastasas	
gggaaacagc		thatasasas		Ligitigadi	eceecgggga	1200
getgtttgat	gagcccatct	ttatcacggt	ggtagactct	cgttctctca	ggacagatgc	1260
tctcctcggg	gagttccgga	tggacgtggg	caccatttac	agagagcccc	ggcacgccta	1320
tctcaggaag	tggctgctgc	tctcagaccc	tgatgacttc	tctactagaa	ccagaggeta	1380
cctgaaaaca	agcctttgtg	tactagaacc	tanagacasa	acacatataa	30303333000	1440
cccctctcaa	agesseages	2024402220	caacatacta	gegeeeeegg	agagaaaaga	
CCCCCCCgaa	gacaaggagg	acattgaaag	caacetgete	eggeceaeag	gcgtagccct	1500
gcgaggagcc	cacttctgcc	tgaaggtctt	ccgggccgag	gacttgccgc	agatggacga	1560
tgccgtgatg	gacaacgtga	aacagatctt	tggcttcgag	agtaacaaga	agaacttggt	1620
ggaccccttt	gtggaggtca	actttacaga	gaaaatgctg	tgcagcaaga	tettagagaa	1680
gacggccaac	cctcagtgga	accadaacat	cacactocct	accatattta	cotoggagaa	
gaoggoodac	occount and a	testesses	cacactgcct	gccacgccc	ceceatgeg	1740
cyaaaaaacy	aggattcgta	ccatagactg	ggaccgcctg	actcacaatg	acategtgge	1800
taccacctac	ctgagtatgt	cgaaaatctc	tgcccctgga	ggagaaatag	aagaggagcc	1860
tgcaggtgct	gtcaagcctt	cgaaagcctc	agacttggat	gactacctgg	acttectece	1920
cacttttagg	ccctgctaca	tcaacctcta	tagcagtece	agagagttca	cangetteec	1980
agacccctac	acagagetea	acacacacaa	aggaggaga	atagagtota	ctaggettett	
teteetetee	atasasasas	acacaggcaa	ggggaaggt	gragoriaco	grageeager	2040
Lougettetee	ctggagacca	agetggtgga	geacagegaa	cagaaggtgg	aggaccttcc	2100
tgcggatgac	atcctccggg	tggagaagta	ccttaggagg	cgcaagtact	ccctgtttgc	2160
ggccttctac	tcagccacca	tgctgcagga	tgtggatgat	gccatccagt	ttgaggtcag	2220
catcoggaac	tacgggaaca	agttcgacat	gacctgcctg	ccactaacct	ccaccactca	2280
gtacageegt	gcagtctttg	acquatacca	ctactactac	ctacceteee	at a seat case	
acctatacta	gtagtatat	actactaca	ccaccaccac	ccaccccggg	graacgrgaa	2340
accegegeg	gtgctgtcat	cctactggga	ggacaccage	catagaatcg	agactcagaa	2400
ccagctgctt	gggattgctg	accggctgga	agctggcctg	gagcaggtcc	acctggccct	2460
gaaggcgcag	tgctccacgg	aggacgtgga	ctcgctggtg	gctcagctga	cogatoacct	2520
catcgcaggc	tgcagccagc	ctctgggtga	catccatgag	acacceteta	ccacccacct	2580
ggaccagtac	ctgtaccagc	tacacaccca	tcacctgag	casatcacta	200000000	
agacatassa	ctcccccc	atasaataaa	tagaaatata	caaaccactg	aggetgeeet	2640
ggccccgaag	ctcggccaca	grgagereee	cgcagetetg	gageaggegg	aggactggct	2700
cctgcgtctg	cgtgccctgg	cagaggagcc	ccagaacagc	ctgccggaca	tcgtcatctg	2760
gatgctgcag	ggagacaagc	gtgtggcata	ccagcgggtg	cccgcccacc	aagtcctctt	2820
ctcccqqcqq	ggtgccaact	actgtggcaa	gaattgtggg	aagctacaga	caatctttct	2880
gaaatatccg	atggagaagg	tacctageae	ccadatacca	atacaastac	agatassact	2940
ataatttaaa	ctctctctc	atraragera	attanagaa	tttaataaa	gggccaagct	
5-5-5-5-5-5	ctctctgtgg	acgagaagga	geccaaccag	cutgotgagg	ggaagetgte	3000
tgtctttgct	gaaacctatg	agaacgagac	taagttggcc	cttgttggga	actggggcac	3060
aacgggcctc	acctacccca	agttttctga	cgtcacgggc	aagatcaagc	tacccaagga	3120
cagcttccgc	ccctcggccg	gctggacctg	ggctggagat	tagttcatat	gtccggagaa	3180
gactctgctc	catgacatgg	acoccootca	cctgagette	atagaagagg	tatttaagaa	3240
ccadacccdd	cttcccggag	acceptaget	ctacatoact	goggaagagg	cgcccgagaa	
ccagacccgg	ctcctcggag	gecageggae	hasabasasa	gacaactaca	ccgatgtgaa	3300
cygygagaag	gtgcttccca	aggacgacac	tgagtgeeca	ctgggctgga	agtgggaaga	3360
tgaggaatgg	tccacagacc	tcaaccgggc	tgtcgatgag	caaggctggg	agtatagcat	3420
caccatcccc	ccggagcgga	agccgaagca	ctgggtccct	gctqaqaaqa	tgtactacac	3480
acaccgacgg	cggcgctggg	tacacctaca	caggagggat	ctcagccaaa	tagaaggagt	3540
gaaaagggag	aggcaggcgg	addcddaddd	caaaaactaa	gagtacgcct	ctettttee	
ctagaagtta	-2222-23	-22-22-233	-3-3390099	Sagencycct		3600
toggaagete	cacctcgagt	accycaagac	ayatyccttc	caccaccacc	gc <b>rggcgcc</b> g	3660
tcgcatggag	ccactggaga	agacggggcc	tgcagctgtg	tttgcccttg	agggggccct	3720
gggcggcgtg	atggatgaca	agagtgaaga	ttccatgtcc	gtctccacct	tgagcttcgg	3780
tgtgaacaga	cccacgattt	cctgcatatt	cgactatogo	aaccoctacc	atctaccete	3840
ctacatotac	caggcccggg	acctaactac	datacasa	gactotttt	ctcataatt	
taggete	toottooto	2002990090	gacggacaag	guccicitit	cigalocota	3900
Lyccategee	teetteetge	accagageea	gaagacggtg	grggrgaaga	acacccttaa	3960
ccccacctgg	gaccagacgc	tcatcttcta	cgagatcgag	atctttggcg	agccggccac	4020
agttgctgag	caaccgccca	gcattgtggt	ggagctgtac	gaccatgaca	cttatggtgc	4080
agacgagttt	atgggtcgct	gcatctgtca	accgagteto	gaacggatgc	cacgactage	4140
			J J	J J J J J J J J J J J J J J J J J J J	- 33 35 -	

```
ctggttccca ctgacgaggg gcagccagcc gtcgggggag ctgctggcct cttttgagct
                                                                                      4200
catccagaga gagaagccgg ccatccacca tattcctggt tttgaggtgc aggagacatc
                                                                                      4260
                                                                                      4320
aaggatectg gatgagtetg aggacacaga cetgecetae ecaceacece agagggagge
caacatetac atggtteete agaacateaa gecagegete eagegtaceg ecategagat ectggcatgg ggcetgegga acatgaagag ttaccagetg gecaacatet ecteecceag
                                                                                      4380
                                                                                      4440
cctcgtggta gagtgtgggg gccagacggt gcagtcctgt gtcatcagga acctccggaa
                                                                                      4500
gaaccccaac tttgacatct gcaccctctt catggaagtg atgctgccca gggaggagct
                                                                                      4560
                                                                                      4620
ctactgccc cccatcaccg tcaaggtcat cgataaccgc cagtttggcc gccggcctgt
ggtgggccag tgtaccatcc gctccctgga gagcttcctg tgtgacccct actcggcgga
                                                                                      4680
gagtccatcc ccacagggtg gcccagacga tgtgagccta ctcagtcctg gggaagacgt gctcatcgac attgatgaca aggagccct catccccatc caggaggaag agttcatcga
                                                                                      4740
                                                                                      4800
ttggtggagc aaattetttg cetecatagg ggagagggaa aagtgegget cetacetgga
                                                                                      4860
gaaggatttt gacaccctga aggtctatga cacacagctg gagaatgtgg aggcctttga gggcctgtct gacttttgta acaccttcaa gctgtaccgg ggcaagacgc aggaggagac
                                                                                      4920
                                                                                      4980
agaagateca tetgtgattg gtgaatttaa gggcetette aaaatttate eeeteecaga agacceagee atececatge eeccaagaca gttecaceag etggeegeee agggaceeca
                                                                                      5040
                                                                                      5100
                                                                                      5160
ggagtgcttg gtccgtatct acattgtccg agcatttggc ctgcagccca aggaccccaa
                                                                                      5220
tggaaagtgt gatccttaca tcaagatctc catagggaag aaatcagtga gtgaccagga
taactacate ceetgeacge tggagecegt atttggaaag atgttegage tgacetgeac
                                                                                      5280
tetgeetetg gagaaggace taaagateae tetetatgae tatgacetee tetecaagga
                                                                                      5340
cgaaaagate ggtgagacgg tegtegacet ggagaacagg etgetgteca agtttgggge
                                                                                      5400
tegetgtgga eteceacaga cetactgtgt etetggaceg aaccagtgge gggaccaget
                                                                                      5460
ecgecetee cageteetee acetettetg ccageageat agagteaagg cacetgtgta
                                                                                      5520
ccggacagac cgtgtaatgt ttcaggataa agaatattcc attgaagaga tagaggctgg
                                                                                      5580
caggatecca aacccacace tgggcccagt ggaggagegt etggetetge atgtgettea
                                                                                      5640
gcagcagggc ctggtcccgg agcacgtgga gtcacggccc ctctacagcc ccctgcagcc agacatcgag caggggaagc tgcagatgtg ggtcgaccta tttccgaagg ccctggggcg
                                                                                      5700
                                                                                      5760
gcctggacct cccttcaaca tcaccccacg gagagccaga aggtttttcc tgcgttgtat
                                                                                      5820
tatetggaat accagagatg tgateetgga tgacetgage etcaeggggg agaagatgag
                                                                                      5880
cgacatttat gtgaaaggtt ggatgattgg ctttgaagaa cacaagcaaa agacagacgt
                                                                                      5940
gcattategt tecetgggag gtgaageaac tteaactgga ggtteatttt ececttegae
                                                                                      6000
tacctgccag ctgagcaagt ctgtaccatt gccaagaagg atgccttctg gaggctggac aagactgaga gcaaaatccc agcacgagtg gtgttccaga tctgggacaa tgacaagttc
                                                                                      6060
                                                                                      6120
tcetttgatg attttetggg cteectgeag etegatetea acegeatgee caagecagee aagacageca agaagtgete ettggaccag etggatgatg etttecacce agaatggtt
                                                                                      6180
                                                                                      6240
                                                                                      6300
gtgtcccttt ttgagcagaa aacagtgaag ggctggtggc cctgtgtagc agaagagggt
gagaagaaaa tactggcggg caagctggaa atgaccttgg agattgtagc agagagtgag
                                                                                      6360
catgaggage ggeetgetgg ceagggeegg gatgageeca acatgaacee taagettgag
                                                                                      6420
gacccaagge gecegacae etectteetg tggtttacet ecceatacaa gaccatgaag tteateetgt ggeggegttt eeggtgggee ateateetet teateateet etteateetg
                                                                                      6480
                                                                                      6540
ctgctgttcc tggccatctt catctacgcc ttcccgaact atgctgccat gaagctggtg
                                                                                      6600
 aagcccttca gctgaggact ctcctgccct gtagaagggg ccgtggggtc ccctccagca
                                                                                      6660
tgggactgge etgeeteete egeccagete ggegagetee tecagacete etaggeetga
                                                                                      6720
                                                                                      6780
 ttgtcctgcc agggtgggca gacagacaga tggaccggcc cacactccca gagttgctaa
catggagete tgagateace ceaetteeat cattteette teccecaace caacgetttt ttggateage teagacatat tteagtataa aacagttgga accacaaaaa aaaaaaaaa
                                                                                      6840
                                                                                      6900
                                                                                      6910
 aaaaaaaaa
```

<210> 16 <211> 6911

<212> DNA

<213> Homo sapiens

#### <400> 16

tcgaccgccc agccaggtgc aaaatgccgt gtcattggga gactccgcag ccggagcatt 60 agattacago togacggago togggaaggg oggogggggt ggaagatgag cagaagcooc 120 180 gcccactgga gcagccgggg gtggcccgtt cccctttaag agcaactgct ctaagccagg agccagagat tcgagccggc ctcgcccagc cagccctctc cagcgagggg acccacaagc ggcgcctcgg ccctcccgac ctttccgagc cctctttgcg ccctgggcgc acggggccct 240 300 360 acacgogoca agcatgotga gggtottoat cototatgoo gagaacgtoo acacaccoga 420 480 caccgacate agegatgeet actgeteege ggtgtttgea ggggtgaaga agagaaceaa agtcatcaag aacagcgtga accetgtatg gaatgaggga titgaatggg accecaaggg catcccctg gaccaggget ctgagettea tgtggtggte aaagaccatg agacgatggg 540 600 gaggaacagg ttcctggggg aagccaaggt cccactccga gaggtcctcg ccacccttag tctgtccgcc agcttcaatg ccccctgct ggacaccaag aagcagccca caggggcctc 660 720 getggteetg caggtgteet acacaceget geetggaget gtgeecetgt teeegeeeee 780

tactcctctg gagccctccc cgactctgcc tgacctggat gtagtggcag acacaggagg 840 900 agaggaagac acagaggacc agggactcac tggagatgag gcggagccat tcctggatca aagcggagge cegggggete ceaceacece aaggaaacta cetteacgte etecgececa 960 ctacccggg atcaaaagaa agcgaagtgc gcctacatct agaaagctgc tgtcagacaa accgcaggat ttccagatca gggtccaggt gatcgagggg cgccagctgc cgggggtgaa 1020 1080 catcaagcet gtggtcaagg ttaccgetge agggcagace aageggaege ggatecacaa 1140 gggaaacagc ccactcttca atgagactct tttcttcaac ttgtttgact ctcctgggga 1200 gctgtttgat gagcccatct ttatcacggt ggtagactct cgttctctca ggacagatgc tctcctcggg gagttccgga tggacgtggg caccatttac agagagcccc ggcacgccta tctcaggaag tggctgctgc tctcagaccc tgatgacttc tctgctgggg ccagaggcta 1260 1320 1380 cctgaaaaca agcctttgtg tgctggggcc tggggacgaa gcgcctctgg agagaaaaga 1440 1500 ccctctgaa gacaaggagg acattgaaag caacctgctc cggcccacag gcgtagccct gcgaggagcc cacttctgcc tgaaggtctt ccgggccgag gacttgccgc agatggacga tgccgtgatg gacaacgtga aacagatctt tggcttcgag agtaacaaga agaacttggt ggaccccttt gtggaggtca gctttgcggg gaaaatgctg tgcagcaaga tcttggagaa gacggccaac cctcagtgga accagaacat cacactgcct gccatgtttc cctccatgtg 1560 1620 1680 1740 cgaaaaaatg aggattcgta tcatagactg ggaccgcctg actcacaatg acatcgtggc taccacctac ctgagtatgt cgaaaatctc tgcccctgga ggagaaatag aagaggagcc tgcaggtgct gtcaagcctt cgaaagcctc agacttggat gactacctgg gcttcctccc 1800 1860 1920 cacttttggg ccctgctaca tcaacctcta tggcagtccc agagagttca caggcttccc 1980 2040 agacccctac acagagetea acacaggeaa gggggaaggt gtggettate gtggeegget 2100 tetgetetee etggagacca agetggtgga geacagtgaa cagaaggtgg aggacettee tgcggatgac atcctccggg tggagaagta ccttaggagg cgcaagtact ccctgtttgc ggccttctac tcagccacca tgctgcagga tgtggatgat gccatccagt ttgaggtcag 2160 2220 categggaac tacgggaaca agttegacat gacetgeetg cegetggeet ceaceactea 2280 gtacagecgt geagtetttg acgggtgeca etactactae etaccetggg gtaacgtgaa 2340 2400 acctgtggtg gtgctgtcat cctactggga ggacatcagc catagaatcg agactcagaa 2460 ccagctgctt gggattgctg accggctgga agctggcctg gagcaggtcc acctggccct gaaggegeag tgeteeaegg aggaegtgga etegetggtg geteagetga eggatgaget categeagge tgeageeage etetgggtga cateeatgag acaecetetg ceaceeaect 2520 2580 ggaccagtac ctgtaccage tgegcaccea teacetgage caaateactg aggetgeeet 2640 2700 ggccctgaag ctcggccaca gtgagctccc tgcagctctg gagcaggcgg aggactggct cctgcgtctg cgtgccctgg cagaggagcc ccagaacagc ctgccggaca tcgtcatctg 2760 gatgctgcag ggagacaagc gtgtggcata ccagcgggtg cccgcccacc aagtcctctt ctcccggcgg ggtgccaact actgtggcaa gaattgtggg aagctacaga caatctttct gaaatatccg atggagaagg tgcctggcgc ccggatgcca gtgcagatac gggtcaagct 2820 2880 2940 gtggtttggg ctctctgtgg atgagaagga gttcaaccag tttgctgagg ggaagctgtc tgtctttgct gaaacctatg agaacgagac taagttggcc cttgttggga actggggcac 3000 3060 3120 aacgggcctc acctacccca agttttctga cgtcacgggc aagatcaagc tacccaagga cagetteege cecteggeeg getggaeetg ggetggagat tggttegtgt gteeggagaa gaetetgete catgaeatgg acgeeggtea cetgagette gtggaagagg tgtttgagaa 3180 3240 3300 ccagacccgg cttcccggag gccagtggat ctacatgagt gacaactaca ccgatgtgaa 3360 cggggagaag gtgcttccca aggatgacat tgagtgccca ctgggctgga agtgggaaga tgaggaatgg tccacagacc tcaaccgggc tgtcgatgag caaggctggg agtatagcat caccatccc ccggagcgga agccgaagca ctgggtccct gctgagaaga tgtactacac 3420 3480 acaccgaegg eggegetggg tgegeetgeg caggagggat eteagecaaa tggaagcaet gaaaaggeac aggeaggegg aggeggaggg egagggetgg gagtaegeet etetttttgg etggaagtte cacctegagt accgeaagae agatgeette egeegeegee getggegeeg 3540 3600 3660 tegeatggag ceaetggaga agaeggggee tgeagetgtg titgeeettg agggggeeet gggeggegtg atggatgaca agagtgaaga tteeatgtee gteteeacet tgagettegg 3720 3780 tgtgaacaga cccacgattt cctgcatatt cgactatggg aaccgctacc atctacgctg 3840 ctacatgtac caggcccggg acctggctgc gatggacaag gactcttttt ctgatcccta 3900 tgccatcgtc tccttcctgc accagagcca gaagacggtg gtggtgaaga acacccttaa 3960 ccccacctgg gaccagacgc tcatcttcta cgagatcgag atctttggcg agccggccac 4020 agttgctgag caaccgcca gcattgtggt ggagctgtac gaccatgaca cttatggtgc agacgagttt atgggtcgct gcatctgtca accgagtctg gaacggatgc cacggctggc 4080 4140 ctggttccca ctgacgaggg gcagccagcc gtcgggggag ctgctggcct cttttgagct catccagaga gagaagccgg ccatccacca tattcctggt tttgaggtgc aggagacatc 4200 4260 aaggateetg gatgagtetg aggacacaga cetgeeetae eeaccaceee agagggagge 4320 caacatctac atggttcctc agaacatcaa gccagcgctc cagcgtaccg ccatcgagat 4380 cctggcatgg ggcctgcgga acatgaagag ttaccagctg gccaacatct cctccccag 4440 4500 cctcgtggta gagtgtgggg gccagacggt gcagtcctgt gtcatcagga acctccggaa gaaccccaac tttgacatct gcaccctctt catggaagtg atgctgccca gggaggagct 4560 ctactgcccc cccatcaccg tcaaggtcat cgataaccgc cagtttggcc gccggcctgt 4620 ggtgggccag tgtaccatcc gctccctgga gagcttcctg tgtgacccct actcggcgga gagtccatcc ccacagggtg gcccagacga tgtgagccta ctcagtcctg gggaagacgt 4680 4740 4800 gctcatcgac attgatgaca aggagcccct catccccatc caggaggaag agttcatcga

		cotcoat and	прадародаа	aagtgcggct	cctacctgga	4860
ttggtggagc	aaattctttg	acctetates	cacacageta	gagaatgtgg	aggcctttga	4920
gaaggatttt	gacaccctga	aggeetteaa	actataccaa	ggcaagacgc	aggaggagac	4980
gggcctgtct	gactttgta	at another a	gctgtaccgg	aaaatttatc	••. • •	5040
agaagatcca	tctgtgattg	gcgaacceaa	gttccaccag	ctggccgccc	agggacccca	5100
agacccagcc	atccccatgc	ceceaayaca	agcatttggc	ctgcagccca		5160
ggagtgcttg	gtccgtatct	tesaccitecte		aaatcagtga		5220
tggaaagtgt	gatccttaca	teaagacece	atttggaaag		tgacctgcac	5280
taactacatc	ccctgcacgc	tagageeege			tctccaagga	5340
tctgcctctg	gagaaggacc	taaagateae	ccccacgac	ctactataca	agtttggggc	5400
cgaaaagatc	ggtgagacgg	tegtegaeet	gyayaacagg	aaccagtggc	gggaccagct	5460
tcgctgtgga	ctcccacaga	cctactgtgt	ctctggaccg	agagtcaagg		5520
ccgcccctcc	cagctcctcc	acctettetg	ccagcagcat	agageeaagg	tagaggctgg	5580
ccggacagac	cgtgtaatgt	ttcaggataa	agaatattcc		atgtgcttca	5640
caggatccca	aacccacacc	tgggcccagt	ggaggagcgt	etetage	coctaceacc	5700
gcagcagggc	ctggtcccgg	agcacgtgga	gtcacggccc	tttacagec	ccctgcagcc	5760
	Caddadaadc	tgcagatgtg	ggtcgaccta	ttteegaagg	teesttetat	5820
gcctggacct	cccttcaaca	tcaccccacg	gagagccaga	aggettetee	agaagatgag	5880
tatctggaat	accagagatg	tgatcctgga	tgacctgagc	Ctcacggggt	agaagacgag	5940
cgacatttat	gtgaaaggtt	agatgattgg	ctttgaagaa	cacaagcaaa	toccettees	6000
gcattatcgt	tccctgggag	gtgaaggcaa	cttcaactgg	aggtteattt		6060
	actaaacaaa	tototaccat	tgccaagaag	gatgccttct	ggaggergga	6120
	agcaaaatcc	cagcacgagt	ggtgttccag	atctgggaca	atgacaagut	6180
	~~++++~+	actecetaca	actcaatctc	aaccycacyc	ccaagecage	6240
caagacagcc	aagaagtgct	ccttggacca	gctggatgat	gettteeace	cagaacggcc	6300
tatateeett	tttgagcaga	aaacagtgaa	qqqctqqtqq	Coccycycag	cagaagaggg	
+~~~~~~	atactoocoo	acaaactaaa	aatqacctig	gagattgtag	cagagagtga	6360
gratgaggag	caacctacta	gccagggccg	qqatqagccc	aacatgaacc	ctaagcttga	6420
goacgaggg	cgccccgaca	cctccttcct	gtggtttacc	tececataca	agaccacgaa	6480
gttcatcctg	tggcggcgtt	tccqqtqqqc	catcatcctc	ttcatcatcc	tcttcatcct	6540
geteatette	ctggccatct	tcatctacgc	cttcccgaac	tatgctgcca	tgaagctggt	6600
geogeogee	anctraggac	tetectacce	tgtagaaggg	gccgtggggt	cccctccagc	6660
	actacctcct	ccacccaact	caacaaactc	ctccagacct	cctaggcctg	6720
atgggactgg	caggatagac	agacagacag	atggaccggc	CCACACTCCC	agagttgcta	6780
	- ctcacatcac	cccacttcca	tcatttcct	ULUCUCAAC	ccaacgooo	6840
acatggagee	ctcagacoto	tttcagtata	aaacagttgg	aaccacaaaa	aaaaaaaaa	6900
		- ccccageas-				6911
aaaaaaaaa	. а					

<210> 17 <211> 6911 <212> DNA <213> Homo sapiens

tegacegeee agecaggtge aaaatgeegt gteattggga gaeteegeag eeggageatt 60 agattacage tegaeggage tegggaaggg eggegggggt ggaagatgag eagaageeee tgttetegga aegeeggetg acaagegggg tgagegeagg eggggegggg acceageeta 120 180 gcccactgga gcagccgggg gtggcccgtt cccctttaag agcaactgct ctaagccagg 240 agccagagat tegageegge etegeceage cageeetete cagegagggg acceacaage 300 360 ggcgcctcgg ccctcccgac ctttccgagc cctctttgcg ccctgggcgc acggggccct acacgegeca ageatgetga gggtetteat cetetatgee gagaacgtee acacacega cacegacate agegatgeet actgeteege ggtgtttgea ggggtgaaga agagaaccaa 420 480 agtcatcaag aacagcgtga accetgtatg gaatgaggga titgaatggg accecaaggg 540 catcccctg gaccaggget ctgagettea tgtggtggte aaagaccatg agacgatggg 600 gaggaacagg ttcctggggg aagccaaggt cccactccga gaggtcctcg ccacccctag tctgtccgc agcttcaatg ccccctgct ggacaccaag aagcagccca caggggcctc gctggtcctg caggtgtcct acacaccgct gcctggagct gtgcccctgt tcccgccccc 660 720 780 tactcctctg gagccctccc cgactctgcc tgacctggat gtagtggcag acacaggagg agaggaagac acagaggacc agggactcac tggagatgag gcggagccat tcctggatca 840 900 aageggagge cegggggete ceaceacece aaggaaacta cetteaegte etcegecea 960 ctaccccggg atcaaaagaa agcgaagtgc gcctacatct agaaagctgc tgtcagacaa 1020 accgcaggat ttccagatca gggtccaggt gatcgagggg cgccagctgc cgggggtgaa 1080 catcaageet gtggteaagg ttaccgetge agggeagace aageggaege ggatecacaa 1140 gggaaacage ceactettea atgagactet titetteaac tigittgaet etectgggga getgtttgat gageceatet tiateaeggt ggtagaetet egitetetea ggaeagatge 1200 1260 1320 tetecteggg gagtteegga tggacgtggg caccatttae agagageece ggcacgeeta teteaggaag tggetgetge teteagaece tgatgaette tetgetgggg ceagaggeta 1380 cctgaaaaca agcctttgtg tgctggggcc tggggacgaa gcgcctctgg agagaaaaga 1440

ccctctgaa gacaaggagg acattgaaag caacctgctc cggcccacag gcgtagccct gcgaggagcc cacttctgcc tgaaggtctt ccgggccgag gacttgccgc agatggacga tgccgtgatg gacaacgtga aacagatctt tggcttcgag agtaacaaga agaacttggt ggaccccttt gtggaggtca gctttgcgg gaaaatgctg tgcagcaaga tcttggagaa gacagaacat cacagagcct gccatgttc cctccatgtg cgaaaaaatg aggattcgta tcatagactg ggaccgcctg actcacaatg acatcgtggc taccacctac ctgagtatgt cgaaaatctc tgcccctgga ggagaaatag aagaggagcc tgcaggtgct gtcaagcctt cgaaagcctc agacttggat gactacctgg gcttcctccc cacttttggg ccctgctaca tcaacctcta tggcagtccc agagagttca caggcttccc agacccctac acagagetea acacaggeaa gggggaaggt gtggettate gtggeegget tetgetetee etggagacca agetggtgga geacagtgaa cagaaggtgg aggacettee tgcggatgac atcctccggg tggagaagta ccttaggagg cgcaagtact ccctgtttgc ggccttctac tcagccacca tgctgcagga tgtggatgat gccatccagt ttgaggtcag categggaac tacgggaaca agttegacat gacetgeetg cegetggeet ceaccactea gtacagccgt gcagtctttg acgggtgcca ctactactac ctaccctggg gtaacgtgaa acctgtggtg gtgctgtcat cctactggga ggacatcagc catagaatcg agactcagaa ccagetgett gggattgetg accggetgga agetggeetg gageaggtee acctggeett gageaggteg tgetecaegg aggaegtgga etegetggt geteagetga eggatgaget eategeagge tgeageeage etetgggtga cateeatgag acaccetetg ceaeceaect ggaecagtae etgtaceage tgegeaecea teaectgage caaateaetg aggetgeeet ggccctgaag ctcggccaca gtgagctccc tgcagctctg gagcaggcgg aggactggct cctgcgtctg cgtgccctgg cagaggagcc ccagaacagc ctgccggaca tcgtcatctg gatgetgeag ggagacaage gtgtggeata ccagegggtg ccegeceace aagteetett eteceggegg gtgecaact actgtggea gaattgtggg aagetacaga caatettett gaaatateeg atggagaagg tgeetggege eeggatgeea gtgeagatae gggteaaget gtgttttgg ctetetgtgg atgagaagga gtteaaceag tttgetgag ggaagetgte tgtetttget gaaacetatg agaaegagae taagttggee ettgttggga aetggggeae aacgggcete acctacecca agttttetga cgtcacggge aagatcaage tacccaagga cagettecge eccteggeeg getggacetg ggetggagat tggttegtgt gteeggagaa gaetetgete catgacatgg acgeeggtea cetgagette gtggaagagg tgtttgagaa ccagacccgg cttcccggag gccagtggat ctacatgagt gacaactaca ccgatgtgaa cggggagaag gtgcttccca aggatgacat tgagtgccca ctgggctgga agtgggaaga tgagggaatgg tccacagacc tcaaccgggc tgtcgatgag caaggctggg agtatagcat caccatecee ceggagegga agecgaagea etgggteeet getgagaaga tgtactacae acacegaegg eggegetggg tgegeetgeg eaggagggat eteageeaaa tggaageaet gaaaaggeae aggeaggegg aggeggaggg egagggetgg gagtaegeet etettttgg etggaagtte eacetegagt acegeaagae agatgeette egeegeegee getggegeeg tegeatggag ceaetggaga agaeggggee tgeagetgtg titgeeettg agggggeeet gggcggcgtg atggatgaca agagtgaaga ttccatgtcc gtctccacct tgagcttcgg tgtgaacaga cccacgattt cctgcatatt cgactatggg aaccgctacc atctacgctg ctacatgtac caggcccggg acctggctgc gatggacaag gactcttttt ctgatccctatgccatcgtc tccttcctgc accagagcca gaagacggtg gtggtgaaga acacccttaa ccccacctgg gaccagacgc tcatcttcta cgagatcgag atctttggcg agccggccac agttgctgag caaccgcca gcattgtggt ggagctgtac gaccatgaca cttatggtgc agacgagttt atgggtcgct gcatctgtca accgagtctg gaacggatgc cacggctggc ctggttccca ctgacgaggg gcagccagcc gtcgggggag ctgctggcct cttttgagct catccagaga gagaagccgg ccatccacca tattcctggt tttgaggtgc aggagacatc aaggateetg gatgagtetg aggacacaga cetgeeetae ecaceacee agagggagge caacatctac atggttecte agaacatcaa gecagegete cagegtaceg ceategagat cctggcatgg ggcctgcgga acatgaagag ttaccagctg gccaacatct cctccccag cctcgtggta gagtgtgggg gccagacggt gcagtcctgt gtcatcagga acctccggaa gaaccccac tttgacatct gcaccctctt catggaagtg atgctgcca gggaggagct ctactgccc cccatcaccg tcaaggtcat cgataaccgc cagtttggcc gccggcctgt ggtgggccag tgtaccatcc gctccctgga gagcttcctg tgtgacccct actcggcgga gagtecatec ccacagggtg geccagaega tgtgageeta eteagteetg gggaagaegt getcategae attgatgaca aggageceet catececate caggaggaag agtteatega ttggtggagc aaattetttg cetecatagg ggagagggaa aagtgegget cetacetgga gaaggattit gacaccctga aggtctatga cacacagctg gagaatgtgg aggcctttga gggcctgtct gacttttgta acaccttcaa gctgtaccgg ggcaagacgc aggaggagac agaagatcca tetgtgattg gtgaatttaa gggeetette aaaatttate eeeteecaga agacccagee atecccatge eeecaagaca gttecaccag etggeegeec agggaececa ggagtgettg gtccgtatet acattgtccg ageatttggc etgcagecca aggaceccaa tggaaagtgt gateettaca teaagatete catagggaag aaatcagtga gtgaceagga taactacatc ccctgcacgc tggagcccgt atttggaaag atgttcgagc tgacctgcac tctgcctctg gagaaggacc taaagatcac tctctatgac tatgacctcc tctccaagga cgaaaagatc ggtgagacgg tcgtcgacct ggagaacagg ctgctgtcca agtttggggc tcgctgtgga ctcccacaga cctactgtgt ctctggaccg aaccagtggc gggaccagct 

		•**				FFOO
ccgccctcc	cagctcctcc	acctcttctg	ccagcagcat	agagtcaagg	cacctgtgta	5520
CCCCACACAC	catataatat	ttcaggataa	agaatattcc	attgaagaga	cagaggeegg	5580
gaggatecea	aacccacacc	taaacccaat	ggaggagcgt	ctggctctgc	acgegeeeea	5640
acsacsaaac	ctaateeega	agcacgtgga	gtcacggccc	ctctacagec	ccctgcagec	5700
agacatcoag	caggggaage	tacagatata	ggtcgaccta	tttccgaagg	ecergagaca	5760
gcctggacct	cccttcaaca	tcaccccacq	gagagccaga	aggtttttcc	tgcgttgtat	5820
tatctggaat	accadadata	tgatcctgga	tgacctgage	ctcacggggg	agaagatgag	5880
cgacatttat	atassaatt	ggatgattgg	ctttgaagaa	cacaaqcaaa	agacagacgt	5940
gcattatcgt	tecataggee	ggaagaaaga	cttcaactgg	aggttčattt	tccccttcga	6000
geattatege	gctgagcaag	tetetaceat	taccaagaag	gatgeettet	ggaggctgga	6060
ctacctgcca	agcaaaatcc	caccaccact	agtattccag	atctgggaca	atgacaagtt	6120
caagactgag	gattttctgg	caycacgage	actcatctc	aaccocatoc	ccaagccagc	6180
ctcctttgat	gattttetgg	gotteccegca	actagatast	actttccacc	cagaatggtt	6240
caagacagcc	aagaagtgct	cottggacca	geeggaegae	ccctatataa	cagaagaggg	6300
tgtgtccctt	tttgagcaga	aaacagtgaa	gggccggcgg	cecegegeag	cadadadtda	6360
tgagaagaaa	atactggcgg	gcaagctgga	aatgaeettg	gagattgtag	ctaaggggggg	6420
gcatgaggag	cggcctgctg	gccagggccg	agatgagece	aacatgaacc	ccaagecega	6480
~~~~~~~~~	CCCCCCCACA	cctccttcct	atagtttacc	tecceataca	agaccargae	6540
attesteeta	taacaacatt	tecaataaac	Catcatcctc	ttcatcatcc	CCCCCCCC	6600
gctgctgttc	ctggccatct	tcatctacqc	cttcccgaac	tatgetgeea	Lyaagetygt	
assaccette.	agetgaggae	tetectacce	tqtaqaaqqq	gccgtggggt	ecectedage	6660
staggertag	cctacctcct	ccacccaact	cggcgagctc	ctccagacct	cctaggcctg	6720
	caggatagac	agacagacag	atogaccooc	CCACACLCCC	ayayıtıyıta	6780
a a a t a a a a a c t	ctranateac	cccacttcca	tcatttcctt	ctcccccaac	ccaacgcttt	6840
tttacqtcc	ctcagacata	tttcagtata	aaacagttgg	aaccacaaaa	aaaaaaaaa	6900
			3 33			6911
aaaaaaaaa	a					

<210> 18 <211> 6911 <212> DNA <213> Homo sapiens

#### <400> 18

60 tegacegece agecaggtge aaaatgeegt gteattggga gaeteegeag eeggageatt agattacage tegacggage tegggaaggg eggegggggt ggaagatgag cagaageece 120 180 240 300 ggcgcctcgg ccctcccgac ctttccgagc cctctttgcg ccctgggcgc acggggccct 360 acacgogoca agcatgotga gggtottoat cototatgoo gagaacgtoo acacaccoga 420 480 caccgacatc agcgatgcct actgctccgc ggtgtttgca ggggtgaaga agagaaccaa agtcatcaag aacagcgtga accetgtatg gaatgaggga tttgaatggg acctcaaggg catcccctg gaccaggget ctgagettca tgtggtggte aaagaccatg agacgatggg gaggaacagg ttcctggggg aagccaaggt cccactccga gaggtcetcg ccacccctag 540 600 660 720 tetgteegee agetteaatg ecceetget ggacaceaag aageageeea caggggeete getggtcetg caggtgtcet acacaceget geetggaget gtgcccetgt tecegecece tactectetg gagccetece egactetgee tgacetggat gtagtggcag acacaggagg agaggaagac acagaggace agggacteae tggagatgag geggagecat teetggatea 780 840 900 aagcggagge ecgggggete ccaccaccce aaggaaacta cetteacgte eteegeecca 960 1020 ctaccccggg atcaaaagaa agcgaagtgc gcctacatct agaaagctgc tgtcagacaa accgcaggat ttccagatca gggtccaggt gatcgagggg cgccagctgc cggggggtgaa catcaagcct gtggtcaagg ttaccgctgc agggcagacc aagcggacgc ggatccacaa gggaaacagc ccactcttca atgagactct tttcttcaac ttgtttgact ctcctgggga 1080 1140 1200 gctgtttgat gagcccatct ttatcacggt ggtagactct cgttctctca ggacagatgc 1260 teteeteggg gagtteegga tggacgtggg caccatttae agagageece ggcaegeeta teteaggaag tggetgetge teteagaece tgatgaette tetgetgggg ceagaggeta ectgaaaaca ageetttgtg tgetggggee tggggaegaa gegeetetgg agagaaaaga 1320 1380 1440 cccctctgaa gacaaggagg acattgaaag caacctgctc cggcccacag gcgtagccct gcgaggagcc cacttctgcc tgaaggtctt ccgggccgag gacttgccgc agatggacga 1500 1560 tgccgtgatg gacaacgtga aacagatett tggettegag agtaacaaga agaacttggt 1620 ggaccccttt gtggaggtca gctttgcggg gaaaatgctg tgcagcaaga tcttggagaa gacggccaac cctcagtgga accagaacat cacactgct gccatgttte cctccatgtg 1680 1740 cgaaaaaatg aggattegta teatagaetg ggaeegeetg acteacaatg acategtgge 1800 taccacctac ctgagtatgt cgaaaatctc tgcccctgga ggagaaatag aagaggagcc 1860 tgcaggtgct gtcaagcctt cgaaagcctc agacttggat gactacctgg gcttcctccc cacttttggg ccctgctaca tcaacctcta tggcagtccc agagagttca caggcttccc 1920 1980 2040 agacccctac acagagetea acacaggeaa gggggaaggt gtggettate gtggeegget tetgetetee etggagacca agetggtgga geacagtgaa cagaaggtgg aggacettee 2100

2160 tgcggatgac atcctccggg tggagaagta ccttaggagg cgcaagtact ccctgtttgc ggccttctac tcagccacca tgctgcagga tgtggatgat gccatccagt ttgaggtcag 2220 categggaac tacgggaaca agttegacat gacetgeetg cegetggeet ceaceactea gtacageegt geagtetttg acgggtgeea ctactactae ctaccetggg gtaacgtgaa 2280 2340 acctgtggtg gtgctgtcat cctactggga ggacatcagc catagaatcg agactcagaa ccagctgctt gggattgctg accggctgga agctggcctg gagcaggtcc acctggccct 2400 2460 gaaggegeag tgctccacgg aggacgtgga ctcgctggtg gctcagctga cggatgagct 2520 categoage tgcagecage ctetgggtga catecatgag acacectetg ccaeccacet ggaccagtac etgtaccage tgcgcaccca tcacetgage caaatcaetg aggetgeeet 2580 2640 ggccctgaag ctcggccaca gtgagctccc tgcagctctg gagcaggcgg aggactggct 2700 cetgegtetg cgtgccetgg cagaggagee ccagaacage etgeeggaca tegteatetg 2760 gatgetgeag ggagacaage gtgtggeata ccagegggtg ccegeceace aagteetett eteceggegg ggtgecaact actgtggeaa gaattgtggg aagetacaga caatetttet gaaatateeg atggagaagg tgeetggege eeggatgeea gtgeagatae gggteaaget gtgtttggg etetetgtgg atgagaagga gtteaaceag tttgetggga aetggggeae tgtetttget gaaacetatg agaacgagae taagttggee ettgttggga aetggggeae 2820 2880 2940 3000 3060 aacgggcctc acctacccca agttttctga cgtcacgggc aagatcaagc tacccaagga 3120 3180 cagettecge cecteggeeg getggacetg ggetggagat tggttegtgt gteeggagaa gactetgete catgacatgg acgeeggtea cetgagette gtggaagagg tgtttgagaa 3240 ccagacccgg cttcccggag gccagtggat ctacatgagt gacaactaca ccgatgtgaa cggggagaag gtgcttccca aggatgacat tgagtgccca ctgggctgga agtgggaaga tgaggaatgg tccacagacc tcaaccgggc tgtcgatgag caaggctggg agtatagcat 3300 3360 3420 caccatecce ceggagega agecgaagea etgggteeet getgagaaga tgtactacae acacegaegg eggegetggg tgegeetgeg eaggagggat eteagecaaa tggaageaet gaaaaggeae aggeaggegg aggeggaggg egagggetgg gagtaegeet etetttttgg etggaagtte cacctegagt accgcaagae agatgeette egeegeegee getggegeeg 3480 3540 3600 3660 tegcatggag ccactggaga agacggggcc tgcagctgtg tttgcccttg agggggccct 3720 3780 gggcggcgtg atggatgaca agagtgaaga ttccatgtcc gtctccacct tgagcttcgg 3840 tgtgaacaga cocacgattt cotgoatatt cgactatggg aaccgctacc atctacgctg ctacatgtac caggeceggg acctggetge gatggacaag gaetetttt etgateeta tgecategte teetteetge accagageca gaagaeggtg gtggtgaaga acaecettaa 3900 3960 cccacctgg gaccagacgc tcatcttcta cgagatcgag atctttggcg agccggccac 4020 4080 agttgctgag caaccgccca gcattgtggt ggagctgtac gaccatgaca cttatggtgc agacgagttt atgggtcgct gcatctgtca accgagtctg gaacggatgc cacggctggc 4140 ctggttccca ctgacgaggg gcagccagcc gtcgggggag ctgctggcct cttttgagct catccagaga gagaagccgg ccatccacca tattcctggt tttgaggtgc aggagacatc 4200 4260 4320 aagggteetg gatgagtetg aggacacaga cetgeeetae ecaceacee agagggagge 4380 caacatctac atggttcctc agaacatcaa gccagcgctc cagcgtaccg ccatcgagat cctggcatgg ggcctgcgga acatgaagag ttaccagctg gccaacatct cctccccag cctcgtggta gagtgtgggg gccagacggt gcagtcctgt gtcatcagga acctccggaa gaaccccaac tttgacatct gcaccctctt catggaagtg atgctgccca gggaggagct 4440 4500 4560 4620 ctactgccc cccatcaccg tcaaggtcat cgataaccgc cagtttggcc gccggcctgt ggtgggccag tgtaccatcc gctccctgga gagcttcctg tgtgacccct actcggcgga 4680 gagtccatcc ccacagggtg gcccagacga tgtgagccta ctcagtcctg gggaagacgt gctcatcgac attgatgaca aggagccct catccccatc caggaggaag agttcatcga 4740 4800 4860 ttggtggage aaattetttg cetecatagg ggagagggaa aagtgegget cetacetgga 4920 quaggatttt gacaccctga aggtctatga cacacagctg gagaatgtgg aggcctttga 4980 gggcctgtct gacttttgta acaccttcaa gctgtaccgg ggcaagacgc aggaggagac agaagatcca totgtgattg gtgaatttaa gggcototto aaaatttato cootcoaga agacccagco atocccatgo coccaagaca gttocaccag otggcogcoc agggacccca 5040 5100 5160 ggagtgcttg gtccgtatct acattgtccg agcatttggc ctgcagccca aggaccccaa tggaaagtgt gatccttaca tcaagatctc catagggaag aaatcagtga gtgaccagga 5220 taactacate cectgeacge tggagecegt atttggaaag atgttegage tgacetgeac tetgeetetg gagaaggace taaagateac tetetatgae tatgacetee tetecaagga 5280 5340 cgaaaagatc ggtgagacgg tcgtcgacct ggagaacagg ctgctgtcca agtttggggc tcgctgtgga ctcccacaga cctactgtgt ctctggaccg aaccagtggc gggaccagct 5400 5460 cegecettee cageteetee acetettetg ceageageat agagteaagg cacetgtgta 5520 ccggacagac cgtgtaatgt ttcaggataa agaatattcc attgaagaga tagaggctgg 5580 5640 caggatecca aacceacace tgggeecagt ggaggagegt etggetetge atgtgettea gcagcaggge ctggtcccgg agcacgtgga gtcacggccc ctctacagcc ccctgcagcc agacatcgag caggggaagc tgcagatgtg ggtcgaccta tttccgaagg ccctggggcg 5700 5760 geetggaeet ecetteaaca teaceceacg gagagecaga aggtttttee tgegttgtat 5820 tatctggaat accagagatg tgatcctgga tgacctgagc ctcacggggg agaagatgag 5880 cgacatttat gtgaaaggtt ggatgattgg ctttgaagaa cacaagcaaa agacagacgt gcattatcgt tccctgggag gtgaaggcaa cttcaactgg aggttcattt tccccttcga ctacctgca gctgagcaag tctgtaccat tgccaagaag gatgccttct ggaggctgga 5940 6000 6060 caagactgag agcaaaatcc cagcacgagt ggtgttccag atctgggaca atgacaagtt 6120

ctcctttgat gattttct	na actecetaca	gctcgatctc	aaccgcatgc	ccaagccagc	6180
caagacagcc aagaagtg	rt ccttggacca	gctggatgat	gctttccacc	cagaatggtt	6240
tgtgtccctt tttgagca		gaactaataa	ccctatataa	cagaagaggg	6300
tgtgtccctt tttgagca	ya aaacaytyaa	gggccggcgg	coccettatea	cagagagtga	6360
tgagaagaaa atactggc	gg gcaagctgga	aatgacettg	gagactgtag	cagagagega	6420
gcatgaggag cggcctgc	tg gccagggccg	ggatgagccc	aacatgaacc	ctaagcttga	
agagggagg cocccoa	ea cctccttcct	gtggtttacc	tccccataca	agaccatgaa	6480
gttcatcctg tggcggcg	t tecaatagae	catcatcctc	ttcatcatcc	tcttcatcct	6540
gctgctgttc ctggccat		cttccccaac	tatoctocca	tgaaggtggt	6600
gctgctgttc ctggccat	et teatetacge	teteeegaae	caegeegeea	cocctccage	6660
gaageeette agetgagg	ac teteetgeee	tgtagaaggg	geegragage	cccccage	6720
ataggactag cotacete	ct ccacccaact	cggcgagctc	ctccagacct	cctaggeetg	
attatactac caggatag	rc agacagacag	atggaccggc	ccacactccc	agagttgcta	6780
acatggaget etgagate	ac cccacttcca	tcatttcctt	ctccccaac	ccaacgettt	6840
tttggatcag ctcagaca	te teterate	aaagagttaa	aaccacaaaa	aaaaaaaaaa	6900
	ta tttcagtata	aaacagccgg	aassasaaaa		6911
aaaaaaaaa a					9711

<210> 19 <211> 6911 <212> DNA <213> Homo sapiens

<400> 19

60 tcgaccgccc agccaggtgc aaaatgccgt gtcattggga gactccgcag ccggagcatt 120 agattacage tegacggage tegggaaggg eggegggggt ggaagatgag eagaageece tgttetegga acgceggetg acaagegggg tgagegeagg eggggegggg acceagecta geceaetgga geageeggg gtggeeegtt eccetttaag ageaaetget etaageeagg agecagagat tegageegge etegeeeage eageeetete eagegagggg acceaeaage 180 240 300 360 ggcgcctcgg ccctcccgac ctttccgagc cctctttgcg ccctgggcgc acggggccct acacgegeca ageatgetga gggtetteat cetetatgee gagaacgtee acacaceega 420 caccgacate agegatgeet actgeteege ggtgtttgea ggggtgaaga agagaaceaa agteateaag aacagegtga accetgtatg gaatgagga tttgaatggg aceteaaggg catececetg gaccaggget etgagettea tgtggtggte aaagaceatg agaegatggg gaggaacagg tteetggggg aagecaaggt eeeecteega gaggteeteg eeaceectag 480 540 600 660 tetgteegee agetteaatg cececetget ggacaceaag aageageeca caggggeete getggteetg caggtgteet acacaceget geetggaget gtgeeeetgt tecegeecee tacteetetg gageecteee egactetgee tgaeetggat gtagtggeag acacaggagg 720 780 840 agaggaagac acagaggacc agggactcac tggagatgag gcggagccat tcctggatca 900 aageggagge cegggggete ceaceacece aaggaaacta cetteaegte eteegeecea 960 ctacccggg atcaaagaa agcgaagtgc gcctacatct agaaagctgc tgtcagacaa 1020 accgcaggat ttccagatca gggtccaggt gatcgagggg cgccagctgc cgggggtgaa catcaagcct gtggtcaagg ttaccgctgc agggcagacc aagcggacgc ggatccacaa 1080 1140 gggaacagc ccactettca atgagactet tttettcaac ttgtttgact eteetgggga 1200 gctgtttgat gagcccatct ttatcacggt ggtagactct cgttctctca ggacagatgc 1260 tetecteggg gagtteegga tggaegtggg caccatttae agagageece ggeaegeeta 1320 teteaggaag tggetgetge teteagacee tgatgactte tetgetgggg ceagaggeta eetgaaaaca ageetttgtg tgetggggee tggggacgaa gegeetetgg agagaaaaga eeeetetgaa gacaaggagg acattgaaag caacetgete eggeecacag gegtageeet gegaggagee caettetgee tgaaggtett eegggeegag gacttgeega agatggaega 1380 1440 1500 1560 tgccgtgatg gacaacgtga aacagatctt tggcttcgag agtaacaaga agaacttggt ggaccccttt gtggaggtca gctttgcggg gaaaatgctg tgcagcaaga tcttggagaa gacggccaac cctcagtgga accagaacat cacactgcct gccatgtttc cctccatgtg 1620 1680 1740 1800 cgaaaaaatg aggattegta teatagaetg ggacegeetg acteacaatg acategtgge taccacctac ctgagtatgt cgaaaatctc tgcccctgga ggagaaatag aagaggagcc 1860 tgcaggtgct gtcaagcctt cgaaagcctc agacttggat gactacctgg gcttcctccc cacttttggg ccctgctaca tcaacctcta tggcagtccc agagagttca caggcttccc 1920 1980 agacccctac acagagetea acacaggeaa gggggaaggt gtggettate gtggeegget 2040 tetgetetee etggagacea agetggtgga geacagtgaa cagaaggtgg aggacettee 2100 tgcggatgac atcctccggg tggagaagta ccttaggagg cgcaagtact ccctgtttgc 2160 2220 ggccttctac tcagccacca tgctgcagga tgtggatgat gccatccagt ttgaggtcag categggaac tacgggaaca agttegacat gacetgeetg eegetggeet ecaceactea gtacageegt geagtetttg acgggtgeea etactactae etaceetggg gtaacgtgaa 2280 2340 acctgtggtg gtgctgtcat cctactggga ggacatcagc catagaatcg agactcagaa 2400 ccagctgctt gggattgctg accggctgga agctggcctg gagcaggtcc acctggccct 2460 gaaggcgcag tgctccacgg aggacgtgga ctcgctggtg gctcagctga cggatgagct catcgcaggc tgcagccagc ctctgggtga catccatgag acacctctg ccaccacct 2520 2580 ggaccagtac ctgtaccage tgegcaccca teacetgage caaateactg aggetgeeet 2640 ggccctgaag ctcggccaca gtgagctccc tgcagctctg gagcaggcgg aggactggct cctgcgtctg cgtgccctgg cagaggagcc ccagaacagc ctgccggaca tcgtcatctg 2700 2760

gatgctgcag ggagacaagc gtgtggcata ccagcgggtg cccgcccacc aagtcctctt 2820 ctcccggcgg ggtgccaact actgtggcaa gaattgtggg aagctacaga caatctttct 2880 2940 gaaatatccg atggagaagg tgcctggcgc ccggatgcca gtgcagatac gggtcaagct gtggtttggg ctctctgtgg atgagaagga gttcaaccag tttgctgagg ggaagctgtc 3000 tgtetttget gaaacetatg agaacgagae taagttggee ettgttggga actggggeae 3060 aacgggcctc acctacccca agttttctga cgtcacgggc aagatcaagc tacccaagga 3120 3180 cagcificege eccteggeeg getggaeetg ggetggagat tggttegtgt gteeggagaa 3240 gactotgoto catgacatgg acgooggtoa cotgagotto gtggaagagg tgtttgagaa ccagacccgg cttcccggag gccagtggat ctacatgagt gacaactaca ccgatgtgaa 3300 cggggagaag gtgcttccca aggatgacat tgagtgccca ctgggctgga agtgggaaga tgaggaatgg tccacagacc tcaaccgggc tgtcgatgag caaggctggg agtatagcat caccatccc ccggagcgga agccgaagca ctgggtccct gctgagaaga tgtactacac 3360 3420 3480 acaccgacgg cggcgctggg tgcgcctgcg caggagggat ctcagccaaa tggaagcact 3540 gaaaaggcac aggcaggcgg aggcggaggg cgagggctgg gagtacgcct ctctttttgg ctggaagttc cacctcgagt accgcaagac agatgccttc cgccgccgcc gctggcgccg tcgcatggag ccactggaga agacggggcc tgcagctgtg tttgcccttg agggggccct 3600 3660 3720 gggcggcgtg atggatgaca agagtgaaga ttccatgtcc gtctccacct tgagcttcgg 3780 tgtgaacaga cccacgattt cctgcatatt cgactatggg aaccgctacc atctacgctg 3840 3900 ctacatgtac caggeceggg acctggetge gatggacaag gactetttt etgateceta tgccatcgtc tccttcctgc accagagcca gaagacggtg gtggtgaaga acacccttaa ccccacctgg gaccagacgc tcatcttcta cgagatcgag atctttggcg agccggccac 3960 4020 4080 agttgctgag caaccgccca gcattgtggt ggagctgtac gaccatgaca cttatggtgc 4140 agacgagttt atgggtcgct gcatctgtca accgagtctg gaacggatgc cacggctggc ctggttccca ctgacgaggg gcagccagcc gtcgggggag ctgctggcct cttttgagct 4200 catccagaga gagaagccgg ccatccacca tattcctggt tttgaggtgc aggagacatc 4260 aaggateetg gatgagtetg aggacacaga cetgecetae ceaceacece agagggagge caacatetae atggtteete agaacateaa gecagegete cagegtaeeg ceategagat 4320 4380 4440 cctggcatgg ggcctgcgga acatgaagag ttaccagctg gccaacatct cctccccag cctcgtggta gagtgtgggg gccagacggt gcagtcctgt gtcatcagga acctccggaa gaaccccaac tttgacatct gcaccctctt catggaagtg atgctgcca gggaggagct ctactgccc cccatcaccg tcaaggtcat cgataaccgc cagtttggcc gccggcctgt ggtgggccag tgtaccatcc gctccctgga gagcttcctg tgtgaccct actcggcgga 4500 4560 4620 4680 4740 gagtccatcc ccacagggtg gcccagacga tgtgagccta ctcagtcctg gggaagacgt 4800 getcategae attgatgaea aggageeeet cateceeate caggaggaag agtteatega ttggtggage aaattetttg cetecatagg ggagagggaa aagtgegget cetacetgga gaaggatttt gacaceetga aggtetatga cacacagetg gagaatgtgg aggeetttga 4860 4920 4980 gggcctgtct gacttttgta acaccttcaa gctgtaccgg ggcaagacgc aggaggagac 5040 agaagatoca totgtgattg gtgaatttaa gggoototto aaaatttato cootoccaga 5100 agacccagce atccccatge ecccaagaca gttecaccag etggeegeec agggaeceea ggagtgcttg gtccgtatct acattgtccg agcatttggc ctgcagccca aggaccccaa tggaaagtgt gatccttaca tcaagatctc catagggaag aaatcagtga gtgaccagga 5160 5220 taactacatc ccctgcacgc tggagcccgt atttggaaag atgttcgagc tgacctgcac 5280 tetgeetetg gagaaggaee taaagateae tetetatgae tatgaeetee tetecaagga 5340 cgaaaagatc ggtgagacgg tcgtcgacct ggagaacagg ctgctgtcca agtttggggc tcgctgtgga ctcccacaga cctactgtgt ctctggaccg aaccagtggc gggaccagct ccgccctcc cagctcctcc acctcttctg ccagcagcat agagtcaagg cacctgtgta 5400 5460 5520 ccggacagac cgtgtaatgt ttcaggataa agaatattcc attgaagaga tagaggctgg 5580 caggatecca aacccacace tgggcccagt ggaggagcgt ctggctctgc atgtgcttca 5640 5700 gcagcaggge ctggteecgg ageacgtgga gteacggeec etetacagee ecetgeagee 5760 agacatcgag caggggaagc tgcagatgtg ggtcgaccta tttccgaagg ccctggggcg gcctggacct cccttcaaca tcaccccacg gagagccaga aggtttttcc tgcgttgtat tatctggaat accagagatg tgatcctgga tgacctgagc ctcacggggg agaagatgag 5820 5880 cgacatttat gtgaaaggtt ggatgattgg ctttgaagaa cacaagcaaa agacagacgt 5940 gcattategt teetggag gtgaaggeaa etteaaetgg aggtteattt teeettega etaeetgea getgageaag tetgtaeeat tgeeaagaag gatgeettet ggaggetgga caagaetgag ageaaaatee eageaegagt ggtgtteeag atetgggaea atgaeaagtt 6000 6060 6120 ctcctttgat gattttctgg gctccctgca gctcgatctc aaccgcatgc ccaagccagc caagacagcc aagaagtgct ccttggacca gctggatgat gctttccacc cagaatggtt 6180 6240 6300 tgtgtccctt tttgagcaga aaacagtgaa gggctggtgg ccctgtgtag cagaagaggg tgagaagaa atactggcgg gcaagctgga aatgaccttg gagattgtag cagagagtga gcatgaggag cgcctgctg gccagggccg ggatgagccc aacatgaacc ctaagcttga ggacccaagg cgccccgaca cctccttcct gtggtttacc tccccataca agaccatgaa 6360 6420 6480 6540 gttcatcctg tggcggtgtt tccggtgggc catcatcctc ttcatcatcc tcttcatcct getgetgtte etggeeatet teatetacge ettecegaae tatgetgeea tgaagetggt 6600 gaageeette agetgaggae teteetgeee tgtagaaggg geegtggggt eeeeteeage atgggaetgg eetgeeteet eegeecaget eggegagete etceagaeet eetaggeetg 6660 6720 attgtcctgc cagggtgggc agacagacag atggaccggc ccacactccc agagttgcta 6780

6840

```
acatggaget etgagateae eccaetteca teattteett eteccecaae ecaaegettt
tttggatcag ctcagacata tttcagtata aaacagttgg aaccacaaaa aaaaaaaaa
                                                                                           6900
                                                                                           6911
aaaaaaaaaa a
         <210> 20
         <211> 6911
         <212> DNA
         <213> Homo sapiens
         <400> 20
                                                                                              60
tegacegece agecaggtge aaaatgeegt gteattggga gacteegeag eeggageatt
agattacage tegacggage tegggaaggg eggegggggt ggaagatgag cagaageece
                                                                                            120
180
                                                                                            240
                                                                                            300
ggcgcctcgg ccctcccgac ctttccgagc cctctttgcg ccctgggcgc acggggccct
                                                                                            360
                                                                                            420
acacgegeca ageatgetga gggtetteat cetetatgee gagaacgtee acacaceega
caccgacate agegatgeet actgeteege ggtgtttgea ggggtgaaga agagaaceaa
                                                                                            480
agtcatcaag aacagcgtga accetgtatg gaatgaggga tttgaatggg acctcaaggg catccccctg gaccagggct ctgagcttca tgtggtggtc aaagaccatg agacgatggg
                                                                                            540
                                                                                            600
gaggaacagg ttcctggggg aagccaaggt cccactccga gaggtcctcg ccaccctag
                                                                                            660
tetgteegee agetteaatg ececetget ggacaccaag aageageeea caggggeete
                                                                                            720
gctggtcctg caggtgtcct acacaccgct gcctggagct gtgcccctgt tcccgccccc tactcctctg gagccctccc cgactctgcc tgacctggat gtagtggcag acacaggagg agaggaagac acagaggaccac tggagatgag gcggagccat tcctggatca
                                                                                            780
                                                                                            840
                                                                                            900
aagcggagge ccgggggete ccaccaccc aaggaaacta ccttcacgte ctccgccca
                                                                                            960
ctaccccggg atcaaaagaa agcgaagtgc gcctacatct agaaagctgc tgtcagacaa accgcaggat ttccagatca gggtccaggt gatcgagggg cgccagctgc cgggggtgaa
                                                                                           1020
                                                                                           1080
                                                                                           1140
catcaageet gtggteaagg ttaccgetge agggeagace aageggaege ggatecacaa
gggaaacagc ccactettea atgagactet tttetteaac ttgtttgaet eteetgggga
                                                                                           1200
gctgtttgat gagcccatct ttatcacggt ggtagactct cgttctctca ggacagatgc
                                                                                           1260
tetecteggg gagtteegga tggacgtggg caccatttae agagageece ggeacgeeta teteaggaag tggetgetge teteagacee tgatgaette tetgetgggg ecagaggeta
                                                                                           1320
                                                                                           1380
                                                                                           1440
cctgaaaaca agcctttgtg tgctggggcc tggggacgaa gcgcctctgg agagaaaaga
ccctctgaa gacaaggagg acattgaaag caacctgctc cggcccacag gcgtagccct gcgaggagcc cacttctgcc tgaaggtctt ccgggccgag gacttgccgc agatggacga
                                                                                            1500
                                                                                            1560
tgccgtgatg gacaacgtga aacagatctt tggcttcgag agtaacaaga agaacttggt ggaccccttt gtggaggtca gctttgcggg gaaaatgctg tgcagcaaga tcttggagaa
                                                                                            1620
                                                                                            1680
gacggccaac cctcagtgga accagaacat cacactgcct gccatgtttc cctccatgtg
                                                                                            1740
cgaaaaaatg aggattegta teatagaetg ggaeegeetg acteacaatg acategtgge taccaectae etgagtatgt egaaaatete tgeeeetgga ggagaaatag aagaggagee
                                                                                            1800
                                                                                            1860
tgcaggtgct gtcaagcctt cgaaagcctc agacttggat gactacctgg gcttcctccc
                                                                                            1920
 cactitigg coctgetaca teaaceteta tggcagtece agagagttea caggettece
                                                                                            1980
                                                                                            2040
 agacccctac acagagetea acacaggeaa gggggaaggt gtggettate gtggeegget
 tetgetetee etggagacea agetggtgga geacagtgaa cagaaggtgg aggacettee
                                                                                            2100
 tgcggatgac atcctccggg tggagaagta ccttaggagg cgcaagtact ccctgtttgc ggccttctac tcagccacca tgctgcagga tgtggatgat gccatccagt ttgaggtcag
                                                                                            2160
                                                                                            2220
 catcgggaac tacgggaaca agttcgacat gacctgcctg ccgctggcct ccaccactca
                                                                                            2280
 gtacagecgt geagtetttg acgggtgeca etactactae etaccetggg gtaacgtgaa
                                                                                            2340
 acctgtggtg gtgctgtcat cctactggga ggacatcagc catagaatcg agactcagaa
                                                                                            2400
 ccagetgett gggattgetg accggetgga agetggeetg gageaggtee acetggeet gaaggegeag tgetecaegg aggaegtgga etegetggtg geteagetga eggatgaget categeagge tgeageeage etetgggtga catecatgag acaccetetg ecaeceaect
                                                                                            2460
                                                                                            2520
                                                                                            2580
 ggaccagtac ctgtaccage tgegeaceca teacetgage caaateactg aggetgeeet
                                                                                            2640
 ggccctgaag ctcggccaca gtgagctccc tgcagctctg gagcaggcgg aggactggct
                                                                                            2700
 ectgegtetg egtgeectgg cagaggagee ecagaacage etgeeggaca tegteatetg
                                                                                            2760
 gatgetgeag ggagacaage gtgtggeata ccagegggtg ccegeceace aagteetett
                                                                                            2820
 ctcccggcgg ggtgccaact actgtggcaa gaattgtggg aagctacaga caatctttct
                                                                                            2880
 gaaatateeg atggagaagg tgeetggege eeggatgeea gtgeagatae gggteaaget
                                                                                            2940
 gtggtttggg ctctctgtgg atgagaagga gttcaaccag tttgctgagg ggaagctgtc tgtctttgct gaaacctatg agaacgagac taagttggcc cttgttggga actggggcac aacgggcctc acctaccca agttttctga cgtcacgggc aagatcaagc tacccaagga
                                                                                            3000
                                                                                            3060
                                                                                            3120
 cagetteege eccteggeeg getggaeetg ggetggagat tggttegtgt gteeggagaa
                                                                                            3180
 gactetgete catgacatgg acgeeggtea cetgagette gtggaagagg tgtttgagaa
                                                                                            3240
 ccagacccgg cttcccggag gccagtggat ctacatgagt gacaactaca ccgatgtgaa cggggagaag gtgcttccca aggatgacat tgagtgccca ctgggctgga agtgggaaga tgaggaatgg tccacagacc tcaaccgggc tgtcgatgag caaggctggg agtatagcat
                                                                                            3300
                                                                                            3360
                                                                                            3420
```

```
caccatecee eeggagegga ageegaagea etgggteeet getgagaaga tgtactaeae
                                                                                                                                 3480
acaccgacgg cggcgctggg tgcgcctgcg caggagggat ctcagccaaa tggaagcact
                                                                                                                                 3540
gaaaaggcac aggcaggcgg aggcggaggg cgagggctgg gagtacgcct ctctttttgg ctggaagttc cacctcgagt accgcaagac agatgccttc cgccgccgcc gctggcgccg
                                                                                                                                 3600
                                                                                                                                 3660
tcgcatggag ccactggaga agacggggcc tgcagctgtg tttgcccttg agggggccct
gggcggcgtg atggatgaca agagtgaaga ttccatgtcc gtctccacct tgagcttcgg
                                                                                                                                 3720
                                                                                                                                 3780
                                                                                                                                 3840
tgtgaacaga cccacgattt cctgcatatt cgactatggg aaccgctacc atctacgctg
ctacatgtac caggeceggg acctggetge gatggacaag gactetttt ctgateceta tgccategte teetteetge accagageca gaagacggtg gtggtgaaga acaccettaa eccaecttgg gaccagaege teatetteta egagategag atetttggeg ageeggeeac
                                                                                                                                 3900
                                                                                                                                 3960
                                                                                                                                 4020
                                                                                                                                 4080
agttgctgag caaccgccca gcattgtggt ggagctgtac gaccatgaca cttatggtgc
agacgagttt atgggtcgct gcatctgtca accgagtctg gaacggatgc cacggctggc ctggttccca ctgacgaggg gcagccagcc gtcgggggag ctgctggct cttttgagct catccagga gagaagccgg ccatccacca tattcctggt tttgaggtgc aggagacatc
                                                                                                                                 4140
                                                                                                                                 4200
                                                                                                                                 4260
aaggateetg gatgagtetg aggacacaga cetgecetae ceaceacec agagggagge caacatetae atggtteete agaacateaa gecagegete cagegtaeeg ceategagat
                                                                                                                                 4320
                                                                                                                                 4380
cctggcatgg ggcctgcgga acatgaagag ttaccagctg gccaacatct cctccccag cctcgtggta gagtgtgggg gccagacggt gcagtcctgt gtcatcagga acctccggaa gaaccccaac tttgacatct gcaccctctt catggaagtg atgctgccca gggaggagct ctactgccc cccatcaccg tcaaggtcat cgataaccgc cagtttggcc gccggcctgt
                                                                                                                                 4440
                                                                                                                                 4500
                                                                                                                                 4560
                                                                                                                                 4620
ggtgggccag tgtaccatcc gctccctgga gagcttcctg tgtgacccct actcggcgga
                                                                                                                                 4680
gagtecatec ccacagggtg geccagacga tgtgageeta etcagteetg gggaagacgt
                                                                                                                                 4740
gctcatcgac attgatgaca aggagcccct catccccatc caggaggaag agttcatcga ttggtggagc aaattctttg cctccatagg ggagagggaa aagtgcggct cctacctgga gaaggatttt gacaccctga aggtctatga cacacagctg gagaatgtgg aggcctttga
                                                                                                                                 4800
                                                                                                                                 4860
                                                                                                                                 4920
gggcctgtct gacttttgta acaccttcaa gctgtaccgg ggcaagacgc aggaggagac
                                                                                                                                 4980
                                                                                                                                 5040
agaagatcca tctgtgattg gtgaatttaa gggcctcttc aaaatttatc ccctcccaga
agacccagc atcccatge ccccaagaca gttccaccag ctggccgccc agggacccca
ggagtgcttg gtccgtatct acattgtccg agcatttggc ctgcagccca aggaccccaa
tggaaagtgt gatccttaca tcaagatcte catagggaag aaatcagtga gtgaccagga
                                                                                                                                 5100
                                                                                                                                 5160
                                                                                                                                 5220
                                                                                                                                 5280
 taactacatc ccctgcacgc tggagcccgt atttggaaag atgttcgagc tgacctgcac
 tetgeetetg gagaaggace taaagateae tetetatgae tatgaeetee tetecaagga
                                                                                                                                  5340
cgaaaagatc ggtgagacgg tcgtcgacct ggagaacagg ctgctgtcca agtttggggc tcgctgtgga ctcccacaga cctactgtgt ctctggaccg aaccagtggc gggaccagct ccgccctcc cagctcctcc acctcttctg ccagcagcat agagtcaagg cacctgtgta ccggacagac cgtgtaagt ttcaggataa agaatatcc attgaagaga tagaggctgg
                                                                                                                                 5400
                                                                                                                                  5460
                                                                                                                                  5520
                                                                                                                                  5580
 caggatecca aacccacace tgggcccagt ggaggagegt etggetetge atgtgettea geageaggge etggtecegg ageaegtgga gteaeggeee etetaeagee ecetgeagee agacategag eaggggaage tgeagatgtg ggtegaceta ttteegaagg ecetggggeg
                                                                                                                                  5640
                                                                                                                                  5700
                                                                                                                                  5760
 gcctggacct cccttcaaca tcaccccacg gagagecaga aggtttttcc tgcgttgtat tatctggaat accagagatg tgatcctgga tgacctgage ctcacggggg agaagatgag
                                                                                                                                  5820
                                                                                                                                  5880
 caacattat gtgaaaggtt ggatgattgg ctttgaagaa cacaagcaaa agacagacgt gcgttatcgt tccctggag gtgaaggcaa cttcaactgg aggttcattt tccccttcga gctgagcaag tctgtaccat tgccaagaag gatgccttct ggaggctgga caagactgag agcaaaatcc cagcacgagt ggtgttccag atctggaca atgacaggt ctcctttgat gatttctgg gctccctgca gctggatgat gctgtccag accagcagc caagacagcc aagaagtgct ccttggacca gctggatgat gctttccacc cagaatggtt tgttccact tttgaggaca accagtgaa gctggatgat gctttccacc cagaatggtt
                                                                                                                                  5940
                                                                                                                                  6000
                                                                                                                                  6060
                                                                                                                                  6120
                                                                                                                                  6180
                                                                                                                                  6240
                                                                                                                                  6300
 tgtgtcctt tttgagcaga aaacagtgaa gggctggtgg ccctgtgtag cagaagaggg
 tgagaagaaa atactggcgg gcaagctgga aatgaccttg gagattgtag cagaagagtga gcatgaggag cgcctgctg gccagggccg ggatgagccc aacatgaacc ctaagcttga ggacccaagg cgcccgaca cctccttcct gtggtttacc tccccataca agaccatgaa
                                                                                                                                  6360
                                                                                                                                  6420
                                                                                                                                  6480
 gttcatcctg tggcggcgtt tccggtgggc catcatcctc ttcatcatcc tcttcatcct
                                                                                                                                  6540
 getgetgtte etggeeatet teatetacge ettecegaae tatgetgeea tgaagetggt
                                                                                                                                  6600
 gaagccette agetgaggae teteetgeee tgtagaaggg geegtggggt eeecteeage atgggaetgg cetgeeteet eegeecaget eggegagete etceagaeet eetaggeetg
                                                                                                                                  6660
                                                                                                                                   6720
                                                                                                                                  6780
  attqtcctgc cagggtgggc agacagacag atggaccggc ccacactccc agagttgcta
  acatggaget etgagateac eccaetteca teattteett etecceaac ecaaegettt
                                                                                                                                  6840
  tttggatcag ctcagacata tttcagtata aaacagttgg aaccacaaaa aaaaaaaaa
                                                                                                                                  6900
                                                                                                                                   6911
  aaaaaaaaa a
```

<210> 21

<211> 6909

<212> DNA

<213> Homo sapiens

<400> 21

tcgaccgccc agccaggtgc aaaatgccgt gtcattggga gactccgcag ccggagcatt

120 agattacago togacggago togggaaggg oggogggggt ggaagatgag cagaagooco 180 240 300 ggcgcctcgg ccctcccgac ctttccgagc cctctttgcg ccctgggcgc acggggccct acacgcgcca agcatgctga gggtcttcat cctctatgcc gagaacgtcc acacacccga 360 420 caccgacate agegatgeet actgeteege ggtgtttgea ggggtgaaga agagaaccaa 480 agtcatcaag aacagcgtga accetgtatg gaatgaggga tttgaatggg accetaaggg catcccctg gaccaggget ctgagettea tgtggtggte aaagaccatg agacgatggg gaggaacagg ttcetggggg aagccaaggt cccactccga gaggtcetcg ccacccctag tetgtccgcc agettcaatg cccccctget ggacaccaag aagcagecca caggggcetc 540 600 660 720 gctggtcctg caggtgtcct acacaccgct gcctggagct gtgcccctgt tcccgcccc tactcctctg gagccctccc cgactctgcc tgacctggat gtagtggcag acacaggagg agaggaagac acagaggacc tggagatgag gcggagccat tcctggatca 780 840 900 aageggagge eegggggete ceaceacec aaggaaacta cetteaegte eteegeeca 960 ctaccccggg atcaaagaa agcgaagtgc gcctacatct agaaagctgc tgtcagacaa accgcaggat ttccagatca gggtccaggt gatcgagggg cgccagctgc cgggggtgaa catcaagcct gtggtcaagg ttaccgctgc agggcagacc aagcggacgc ggatccacaa 1020 1080 1140 gggaaacage ccactettea atgagactet titetteaac ttgtitgaet etectgggga 1200 gctgtttgat gagcccatct ttatcacggt ggtagactct cgttctctca ggacagatgc 1260 tetecteggg gagtteegga tggaegtggg caccatttae agagageece ggeaegeeta teteaggaag tggetgetge teteagaece tgatgaette tetgetgggg ceagaggeta 1320 1380 cctgaaaaca agcctttgtg tgctggggcc tggggacgaa gcgcctctgg agagaaaaga 1440 ccctctgaa gacaaggagg acattgaaag caacctgctc cggcccacag gcgtagccct gcgaggagcc cacttctgcc tgaaggtctt ccgggccgag gacttgccgc agatggacga 1500 1560 tgccgtgatg gacaacgtga aacagatctt tggcttcgag agtaacaaga agaacttggt ggaccccttt gtggaggtca gctttgcggg gaaaatgctg tgcagcaaga tcttggagaa gacggccaac cctcagtgga accagaacat cacactgcct gccatgtttc cctccatgtg 1620 1680 1740 cgaaaaatg aggattegta teatagaetg ggacegeetg aeteacaatg acategtgge taccaectae etgagtatgt egaaaatete tgeeeetgga ggagaaatag aagaggagee 1800 1860 tgcaggtgct gtcaagcctt cgaaagcctc agacttggat gactacctgg gcttectccc 1920 cacttttggg ccctgctaca tcaacctcta tggcagtccc agagagttca caggettccc 1980 agaccectae acagagetea acacaggeaa gggggaaggt gtggettate gtggeegget tetgetetee etggagacca agetggtgga gcacagtgaa cagaaggtgg aggacettee 2040 2100 tgeggatgac atecteeggg tggagaagta eettaggagg egeaagtaet eeetgtttge ggeettetae teageeacca tgetgeagga tgtggatgat geeateeagt ttgaggteag 2160 2220 catcgggaac tacgggaaca agttcgacat gacctgcctg ccgctggcct ccaccactca 2280 2340 qtacagccgt gcagtctttg acgggtgcca ctactactac ctaccctggg gtaacgtgaa 2400 acctgtggtg gtgctgtcat cctactggga ggacatcagc catagaatcg agactcagaa ccagetgett gggattgetg accggetgga agetggeetg gageaggtee acetggeeet 2460 gaaggegeag tgetecaegg aggaegtgga etegetggtg geteagetga eggatgaget categeagge tgeageeage etetgggtga categatgag acaecetetg ecaeceaect 2520 2580 2640 ggaccagtac ctgtaccage tgcgcaccca tcacctgage caaatcactg aggetgeect ggecetgaag eteggecaca gtgagetece tgeagetetg gageaggegg aggaetgget eetgegtetg egtgeeetgg eagaggagee ecagaacage etgeeggaca tegteatetg 2700 2760 gatgctgcag ggagacaagc gtgtggcata ccagcgggtg cccgcccacc aagtcctctt 2820 ctcccggcgg ggtgccaact actgtggcaa gaattgtggg aagctacaga caatctttct 2880 2940 gaaatateeg atggagaagg tgeetggege eeggatgeea gtgeagatae gggteaaget gtggtttggg ctctctgtgg atgagaagga gttcaaccag tttgctgagg ggaagctgtc tgtctttgct gaaacctatg agaacgagac taagttggcc cttgttggga actggggcac 3000 3060 3120 aacgggcctc acctacccca agttttctga cgtcacgggc aagatcaagc tacccaagga cagetteege eccteggeeg getggacetg ggetggagat tggttegtgt gteeggagaa 3180 gactetgete catgacatgg acgeeggtea cetgagette gtggaagagg tgtttgagaa ceagaceegg etteeeggag geeagtggat etacatgagt gacaactaca eegatggaa eggggagaag gtgetteeca aggatgacat tgagtgeeca etgggetgga agtgggaaga 3240 3300 3360 tgaggaatgg tccacagacc tcaaccgggc tgtcgatgag caaggctggg agtatagcat 3420 3480 caccatecee eeggagegga ageegaagea etgggteeet getgagaaga tgtactacae 3540 acaccgacgg cggcgctggg tgcgcctgcg caggagggat ctcagccaaa tggaagcact gaaaaggcac aggcaggcgg aggcggaggg cgagggctgg gagtacgcct ctctttttgg ctggaagttc cacctcgagt accgcaagac agatgccttc cgccgccgcc gctggcgccg 3600 3660 tegcatggag ccactggaga agacggggcc tgcagctgtg tttgcccttg agggggccct 3720 gggcggcgtg atggatgaca agagtgaaga ttccatgtcc gtctccacct tgagcttcgg 3780 tgtgaacaga cccacgattt cctgcatatt cgactatggg aaccgctacc atctacgctg 3840 3900 ctacatgtac caggcccggg acctggctgc gatggacaag gactcttttt ctgatcccta tgccatcgtc tccttcctgc accagagcca gaagacggtg gtggtgaaga acacccttaa ccccacctgg gaccagacgc tcatcttcta cgagatcgag atctttggcg agccggccac agttgctgag caaccgccca gcattgtggt ggagctgtac gaccatgaca cttatggtgc 3960 4020 4080

20

#### 32/68

```
4140
agacqaqttt atqqqtcqct gcatctqtca accqaqtctg gaacqqatqc cacqqctqqc
ctggttccca ctgacgaggg gcagccagcc gtcgggggag ctgctggcct cttttgagct catccagaga gagaagccgg ccatccacca tattcctggt tttgaggtgc aggagacatc aaggatcctg gatgagtctg aggacacaga cctgccctac ccaccaccc agagggaggc caacatctac atggttcctc agaacatcaa gccagcgctc cagcgtaccg ccatcgagat
                                                                                                 4200
                                                                                                 4260
                                                                                                 4320
                                                                                                 4380
                                                                                                 4440
cctggcatgg ggcctgcgga acatgaagag ttaccagctg gccaacatct cctccccag
cctcgtggta gagtgtgggg gccagacggt gcagtcctgt gtcatcagga acctccggaa gaaccccaac tttgacatct gcaccctctt catggaagtg atgctgccca gggaggagct
                                                                                                 4500
                                                                                                 4560
ctactgcccc cccatcaccg tcaaggtcat cgataaccgc cagtttggcc gccggcctgt
                                                                                                 4620
ggtgggccag tgtaccatcc gctccctgga gagcttcctg tgtgacccct actcggcgga
                                                                                                 4680
gagtecatec ccacagggtg geccagacga tgtgagecta etcagtectg gggaagacgt getcategae attgatgaea aggageceet catecceate caggaggaag agtteatega
                                                                                                 4740
                                                                                                 4800
ttggtggagc aaattetttg cetecatagg ggagagggaa aagtgegget cetacetgga gaaggatttt gacaceetga aggtetatga cacacagetg gagaatgtgg aggeetttga
                                                                                                 4860
                                                                                                 4920
gggcctgtct gacttttgta acacettcaa getgtaeegg ggeaagaege aggaggagae
                                                                                                 4980
agaagatcca tetgtgattg gtgaatttaa gggcetette aaaatttate eecteecaga agacceagee atceecatge eeceaagaca gtteeaceag etggeegee agggaceeca ggagtgettg gteegtatet acattgteeg ageatttgge etgeageeca aggaceecaa
                                                                                                 5040
                                                                                                 5100
                                                                                                 5160
tggaaagtgt gateettaca teaagatete catagggaag aaateagtga gtgaccagga
                                                                                                 5220
taactacate ceetgeacge tggageeegt atttggaaag atgttegage tgacetgeac
                                                                                                 5280
                                                                                                 5340
tctqcctctg gagaaggacc taaagatcac tctctatgac tatgacctcc tctccaagga
cqaaaagatc ggtgagacgg tcgtcgacct ggagaacagg ctgctgtcca agtttggggc
                                                                                                 5400
tegetgtgga eteceacaga ectaetgtgt etetggaceg aaccagtgge gggaceaget ecgecetec cageteetec acctettetg ecageageat agagteaagg cacetgtgta
                                                                                                 5460
                                                                                                 5520
ccqqacaqac cqtqtaatgt ttcaggataa agaatattcc attgaagaga tagaggctgg
                                                                                                 5580
caggatecca aacccacace tgggcccagt ggaggagcgt ctggctctgc atgtgcttca
                                                                                                 5640
gcagcaggge ctggtcccgg agcacgtgga gtcacggccc ctctacagcc ccctgcagcc agacatcgag caggggaagc tgcagatgtg ggtcgaccta tttccgaagg ccctggggcg
                                                                                                 5700
                                                                                                 5760
                                                                                                 5820
geetggacet cectteaaca teacceeacg gagagecaga aggtttttee tgegttgtat
tatctggaat accagagatg tgatcctgga tgacctgagc ctcacggggg agaagatgag
                                                                                                 5880
                                                                                                 5940
cgacatttat gtgaaaggtt ggatgattgg ctttgaagaa cacaagcaaa agacagacgt
gcattatcgt tccctgggag gtgaaggcaa cttcaactgg aggttcattt tccccttcga ctacctgcca gctgagcaag tctgtaccat tgccaagaag gatgccttct ggaggctgga caagactgag caaaatccca gcacgagtgg tgttccagat ctgggacaat gacaagttct
                                                                                                 6000
                                                                                                 6060
                                                                                                 6120
cetttgatga ttttctgggc tccctgcagc tcgatctcaa ccgcatgccc aagccagcca
                                                                                                 6180
                                                                                                 6240
agacagccaa qaagtgctcc ttggaccagc tggatgatgc tttccaccca gaatggtttg
tgtccctttt tgagcagaaa acagtgaagg gctggtggcc ctgtgtagca gaagagggtg agaagaaaat actggcggc aagctggaaa tgaccttgga gattgtagca gagagtgagc
                                                                                                 6300
                                                                                                 6360
                                                                                                 6420
atgaggagcg gcctgctggc cagggccggg atgagcccaa catgaaccct aagcttgagg
acceaaggeg cecegacace tecticetgt ggtttacete eccatacaag accatgaagt
                                                                                                 6480
tcatcctgtg gcggcgtttc cggtgggcca tcatcctctt catcatcctc ttcatcctgc
                                                                                                 6540
                                                                                                 6600
tgctgttcct ggccatette atctacgect tecegaacta tgctgccatg aagetggtga
agecetteag etgaggaete teetgeeetg tagaagggge egtggggtee cetceageat gggaetggee tgeeteetee geecageteg gegageteet ceagacetee taggeetgat tgteetgeea gggtgggeag acagacagat ggaeeggeee acacteecag agttgetaae
                                                                                                 6660
                                                                                                 6720
                                                                                                 6780
atggagetet gagateacce cacttecate attteettet eccecaacce aacgettttt
                                                                                                 6840
                                                                                                 6900
6909
aaaaaaaa
          <210> 22
          <211> 20
          <212> DNA
          <213> Homo sapiens
          <400> 22
                                                                                                     20
  tgggacctca agggcatccc
          <210> 23
          <211> 20
          <212> DNA
          <213> Homo sapiens
          <400> 23
```

accatgctgc aggatgtgga
<210> 24
<211> 20

<212> DNA <213> Homo sap	iens				
<400> 24 gggaggtgaa ggcaacttc	a				20
<210> 25					
<211> 20					
<212> DNA					
<213> Homo sap	tens				
<400> 25 ctcacggggg agaagatga	g				20
<210> 26					
<211> 20 <212> DNA					
<212> DNA <213> Homo sap	oiens				
<400> 26					
ctgtggcggc gtttccggt	.g				20
<210> 27					
<211> 20 <212> DNA					
<213> Homo sap	oiens				
<400> 27					
acatcaagga tcctggatg	ja –				20
<210> 28					
<211> 20					
<212> DNA					
<213> Homo sar	preus				
<400> 28					20
ctgtggcggc gtttccggt	-g				
<210> 29					
<211> 20					
<212> DNA <213> Homo sa	niens				
	220				
<400> 29 acagacgtgc attatcgtf	tc				20
<210> 30					
<211> 20 <212> DNA					
<213> Homo say	piens				
<400> 30					
aagactgaga gcaaaatc	cc				20
<210> 31					
<211> 507					
<212> DNA	•				
<213> Homo sa	preus				
<400> 31		atanttaa.	aact ccacea	ccaaacatt	60
tcgaccgccc agccaggt agattacagc tcgacgga	ge aaaatgeegt ge tegggaaggg	caacaaaaat	ggaagatgag	cagaagcccc	120
tattetegga acgeegge	to acaageggg	tgagcgcagg	cggggcgggg	acccagecta	180
annestans acaaccaa	aa ataacccatt	cccctttaaq	agcaactgct	ctaagccagg	240
aggregat traggrega	ac ctcacccaac	cagccctctc	caqcqagggg	acccacaage	300 360
ggcgcctcgg ccctcccg	ac ctttccgagc aa aaatcttcat	cctctatgcc	gagaacgtcc	acacacccga	420

caccgacatc agcgatgcct actgct ctcgccgggg tcggggtggg gtagag		ggtaggaggg	gccgaccacc	480 507
<210> 32 <211> 183 <212> DNA <213> Homo sapiens				
<pre>&lt;400&gt; 32 aaaggcggga tgtgtctctc cattct accaaagtca tcaagaacag cgtgaa ttccttttct ctctgtctgc tgcagg ttg</pre>	ccct gtatggaatg	aggtatgtga	gtttttctcc	60 120 180 183
<210> 33 <211> 264 <212> DNA <213> Homo sapiens				
<pre>&lt;400&gt; 33 cattcatgaa tgcctactca gtgccc ctcctagagg gccataggtt aagatg acctcaaggg catcccctg gaccag agacgatggg gaggaacagg taaggt aggtaggatt gtggagtata caga</pre>	cctt ttctctttt ggct ctgagcttca	cttccaggga tgtggtggtc	tttgaatggg aaagaccatg	60 120 180 240 264
<210> 34 <211> 223 <212> DNA <213> Homo sapiens				
<pre>&lt;400&gt; 34 cagaagagcc agggtgcctt aggcta tcctgggga agccaaggtc ccactc gcttcaatgc cccctgctg gacacc ctctgccagg ttaaggtcca aggcat</pre>	cgag aggtcctcgc aaga agcagcccac	cacccctagt aggggtaagt	ctgtccgcca	60 120 180 223
<210> 35 <211> 224 <212> DNA <213> Homo sapiens				
<400> 35 cagtggtccg aggccagcgc accaac tggtcctgca ggtgtcctac acaccg ctcctctgga gccctccccg actctg cgttggcctg gctgggcccc agcaag	ctgc ctggagctgt cctg acctggatgt	gcccctgttc agtggcaggt	ccgcccccta	60 120 180 224
<210> 36 <211> 315 <212> DNA <213> Homo sapiens				
<400> 36 aggggcaggg gcagggccag agggccggaggg agaccagag gaccaggg atcaaagcgg aggcccgggg gctcccccactaccc cgggatcaaa agaaagacaaaccgca ggatttccag gtgatgccatcagctg cgggt	ggac tcactggaga acca ccccaaggaa ggaa gtgcgcctac	tgaggcggag actaccttca atctagaaag	ccattcctgg cgtcctccgc ctqctgtcag	60 120 180 240 300 315
<210> 37 <211> 249 <212> DNA <213> Homo sapiens				

•

<pre>&lt;400&gt; 37 ccagtggtga gatggtccct gagatttctg actcttgggg tggatggtgg gtggtcctt actcttcccc cttctggctt tcagatcagg gtccaggtga tcgaggggcg ccagctgcc ggggtgaaca tcaagcctgt ggtcaaggtt accgctgcag ggcagaccaa gcggacgcg atccacaagg gaaacagccc actcttcaat gaggtgggag acatggggca tgagggcag accttgtgg</pre>	g 120
<210> 38 <211> 185 <212> DNA <213> Homo sapiens	
<400> 38 ccctggcctg agggatcagc aggcactgat atgtctctct ttgctctgaa ccaacagac cttttcttca acttgtttga ctctcctggg gagctgtttg atgagcccat ctttatcac gtatgtctca gcagtcaaag tgttctccgt gggctgtatg tatgcacata ggtgtcagt cacac	g 120
<210> 39 <211> 196 <212> DNA <213> Homo sapiens	
<400> 39 aagagetatt gggttggeeg tgtgggeeae atgteeetgt gaatgtgage catgatett etetgeaggt ggtagaetet egttetetea ggacagatge teteeteggg gagtteegg taattgetta ttttetaaaa geagteagtt eteaettete egtgttggtg gageetetgggaeeatggg eagggg	<u>120</u>
<210> 40 <211> 178 <212> DNA <213> Homo sapiens	
<400> 40  tggaatcgta taatgcacca cactttattt aacgctttgg cggcaagagt ttgatttgt tctcctctct tgattgcaga tggacgtggg caccatttac agagagcccc gtgagttct accactttgg ccgtatcctt gcattttggt tctggaggct gattggggac actcattt	ig 60 ic 120 178
<210> 41 <211> 231 <212> DNA <213> Homo sapiens	
<400> 41 ggggtcttct gattctggga tcaccaaagg atgttgtctc tcttagggca cgcctatct aggaagtggc tgctgctctc agaccctgat gacttctctg ctggggccag aggctacct aaaacaagcc tttgtgtgct ggggcctggg gacgaagcgc ctgtgagtac atttccctg gtcttcctta cggtccccca cgcggcactt ggttgcggag gcaccaaacc a	.g 120
<210> 42 <211> 247 <212> DNA <213> Homo sapiens	
<pre>&lt;400&gt; 42 gtcaaaaccc tgtgctcagg agcgcatgaa ggaacgtatt tggttttett tgtagctgg gagaaaagac ccctctgaag acaaggagga cattgaaagc aacctgctcc ggcccacac cgtagccctg cgaggagccc acttctgcct gaaggtcttc cgggccgagg acttgccg gagtgcgtgg ggcgcccct tgggtgggag gtctgcagga ggctggaggc gcagggctt tgggggt</pre>	gg 120 ca 180
<210> 43 <211> 179 <212> DNA <213> Homo sapiens	

<pre>&lt;400&gt; 43 caggcagtga ctggtgtgtc cctcttccca gtgga cagatctttg gcttcgagag taacaagaag aactt tttgcgggga aaatggtaag gagcaaggga gcagg</pre>	ggtgg acccctttgt ggaggtcagc 120
<210> 44 <211> 202 <212> DNA <213> Homo sapiens	
<pre>&lt;400&gt; 44 ccccggggga gcccagagtc cccatggagc tgatc agctgtgcag caagatcttg gagaagacgg ccaac tgcctgccat ggtgagcctc ctgtccccag caaac cttcgggagg tccagggctc ct</pre>	cctca gtggaaccag aacatcacac 120
<210> 45 <211> 167 <212> DNA <213> Homo sapiens	
<400> 45 gggaggggct gttctatctt caaaaggact cttct agtttccctc catgtgcgaa aaaatgagga ttcgt ttggagtctt tagggcgggc tgtcctgagg gggcg	atcat agactggtga gttctgagtc 120
<210> 46 <211> 220 <212> DNA <213> Homo sapiens	
<pre>&lt;400&gt; 46 tgtggcctga gttcctttcc tgtgtcaggc cctct cctgactcac aatgacatcg tggctaccac ctacc tggaggagaa atagaaggta tgttccctct tcgtt cccccctcta tccagcttac acttctagtt ttgag</pre>	etgagt atgtcgaaaa tetetgeeee 120 eetgee etttgaeeee etgtgetete 180
<210> 47 <211> 172 <212> DNA <213> Homo sapiens	
<pre>&lt;400&gt; 47 acagcctgtt catgtaaccc gtccttctcc cagcc ttctttacgc ttcagaggag cctgcaggtg ctgtc cgttgctgtc accttgggga caaccagggg agtgc</pre>	caagee ttegaaagee teagaetgta 120
<210> 48 <211> 200 <212> DNA <213> Homo sapiens	
<pre>&lt;400&gt; 48 ccgacccctc tgattgccac ttgtgtctcc cagtg cttttgggcc ctgctacatc accetctatg gcagg acccctacac agagctcaac acaggcaagg taagg atgccacatg cccaggtggg</pre>	cccag agagttcaca ggcttcccag 120
<210> 49 <211> 217 <212> DNA <213> Homo sapiens	
<400> 49 cctccctct gtctcccctg ctccttgtga cctgcctatcgtgg ccggcttctg ctctccctgg agacaggtggagga ccttcctgcg gatgacatcc tccgc	caagct ggtggagcac agtgaacaga 120

```
217
gtgggagetg ggegtegggg cagggaaggg atggeca
      <210> 50
      <211> 269
      <212> DNA
      <213> Homo sapiens
      <400> 50
agcctgggtg cctttctttg ctcctcccgt gaccctctgg tctactctct gctctcagaa
                                                                           60
gtaccttagg aggegeaagt acteeetgtt tgeggeette tacteageea écatgetgea
                                                                          120
ggatgtggat gatgccatcc agtttgaggt cagcatcggg aactacggga acaagttcga
                                                                          180
catgacetge etgeegetgg cetecaceae teagtacage egtgeagtet ttgaeggtga
                                                                          240
                                                                          269
ggcagtgctc ctggctggga ccccgatca
      <210> 51
      <211> 225
      <212> DNA
      <213> Homo sapiens
      <400> 51
actoctggca cagogotcag goocgtotot coattocagg gtgccactac tactacotac
                                                                           60
cctggggtaa cgtgaaacct gtggtggtgc tgtcatccta ctgggaggac atcagccata
                                                                          120
                                                                          180
gaatcgagac tcagaaccag ctgcttggga ttgctgaccg gctggtgagt gaaaacttgc
ccaaagctgc acatgcctat gcatgcacct gctacccccg ctgca
                                                                          225
      <210> 52
      <211> 227
      <212> DNA
      <213> Homo sapiens
      <400> 52
gggtccagca tgcaccctct gccctgtggt gacacacctg acccttgcct gcccattcca
                                                                           60
caggaagetg geetggagea ggteeacetg geeetgaagg egeagtgete caeggaggae gtggaetege tggtggetea getgaeggat gageteateg eaggetgeag gtaggggga
                                                                          120
                                                                          180
cctggcgccc ctggtgccca cctctcctgg ctcaactggg cctgttt
                                                                          227
      <210> 53
      <211> 303
      <212> DNA
      <213> Homo sapiens
      <400> 53
tgggagaccc tgggctcatc aggcgcattc catctgtccg tccctcacag ccagcctctg
                                                                           60
ggtgacatec atgagacace ctetgecace cacetggace agtacetgta ceagetgege
                                                                          120
acccatcacc tgagccaaat cactgaggct gccctggccc tgaagctcgg ccacagtgag
                                                                          180
                                                                          240
ctcctgcag ctctggagca ggcggaggac tggctcctgc gtctgcgtgc cctggcagag
gaggtaatta agcctggggg tgcctttctt cttctgctct cctgctgcct ggaacatcag
                                                                          300
                                                                          303
aac
       <210> 54
       <211> 272
       <212> DNA
       <213> Homo sapiens
      <400> 54
cgtgggcctg gtgtgtcacc atccccaccc cgaccaccac cctctgttca gccccagaac
                                                                           60
agectgeegg acategteat etggatgetg cagggagaea agegtgtgge ataccagegg
                                                                          120
gtgcccgccc accaagtect ettetecegg eggggtgcca actaetgtgg caagaattgt
                                                                          180
gggaagctac agacaatett tetgaaagtg agttttettt ttecaagtea tgategtatt
                                                                          240
                                                                          272
tocaacataa ggcctttctc ccatctcttg ct
       <210> 55
       <211> 219
       <212> DNA
```

<213> Homo sapiens

<pre>&lt;400&gt; 55 tgtgggtttc tgtccttctt atgccagtgc agatacgggt aaccagtttg ctgaggggaa ccacctctgc ctcccactac</pre>	caagctgtgg gctgtctgtc	tttgggctct tttgctgaaa	ctgtggatga	gaaggagttc	60 120 180 219
<210> 56 <211> 292 <212> DNA <213> Homo sapie	ns				
<pre>&lt;400&gt; 56 tgcctccac tacctggagc agtatgagaa cgagactaag accccaagtt ttctgacgtc cggccggctg gacctgggct gcagggaggg ctggggagag</pre>	ttggcccttg acgggcaaga qqaqattggt	tcaagctacc tcgtgtgtcc	gggcacaacg caaggacagc ggagaagacg	ggcctcacct ttccgcccct tgagtcgtgg	60 120 180 240 292
<210> 57 <211> 242 <212> DNA <213> Homo sapie	ens				
<pre>&lt;400&gt; 57 tggatggggg cctctccagc ctccatgaca tggacgccgg cggcttcccg gaggccagtg gcactcaggg gcaggtgggg ca</pre>	tcacctgagc gatctacatg	ttcgtggaag agtgacaact	aggtgtttga	gaaccagacc ggtaaagcag	60 120 180 240 242
<210> 58 <211> 215 <212> DNA <213> Homo sapie	ens				
<pre>&lt;400&gt; 58 tcacatctgt ctgtctcctc aaggtgcttc ccaaggatga tggtccacag acctcaaccg gagccccatc cccggcaagc</pre>	cattgagtgc ggctgtcgat	ccactgggct gagcaaggtg	ggaagtggga	agatgaggaa	60 120 180 215
<210> 59 <211> 246 <212> DNA <213> Homo sapie	ens				
<400> 59 agagatggtc ccaggagaga ccccaggctg ggagtatagc ctgctgagaa gatgtactac atctcagcca aatggaagca gcctgt	atcaccatcc acacaccgac	ccccggagcg ggcggcgctg	gaagccgaag ggtgcgcctg	cactgggtcc	60 120 180 240 246
<210> 60 <211> 253 <212> DNA <213> Homo sapie	ens				
<400> 60 cttcccaccg gcctctgagt gcgagggctg ggagtacgcc cagatgcctt ccgccgccgc ctgcagctgt gtttgccctt gggtagggta tat	tctctttttg	gctggaagtt	ccacctcgag gccactggag	taccgcaaga aagacggggc	60 120 180 240 253
<210> 61 <211> 177					

	<212> DN <213> Ho	A mo sapie	ens				
gaatct	atcc atc	ccagctt	cgtgtctcca tgagcttcgg ggcctctatg	tgtgaacaga	cccacgattt	cctgcatatt	60 120 177
•	<210> 62 <211> 18 <212> DN <213> Ho	1	ens				
aaccac	tota coc	actcact	ctggcacctc tgtaccaggc gaggcaggag	ccgggacctg	gctgcgatgg	acaaggactc	60 120 180 181
	<210> 63 <211> 31 <212> DN <213> Ho	.9	ens				
ccccac ccctct gtggtg gagatc tacgac	ggcc atog gtga aga tttg gco	tggagaa gcagatcc acaccct gagccggc acttatgt	gacatetete etatgecate taaceceace cacagttget gagtetgece	gtctccttcc tgggaccaga gagcaaccgc	cgctcatctt ccagcattgt	ccagaagacg ctacgagatc ggtggagctg	60 120 180 240 300 319
	<210> 64 <211> 24 <212> DN <213> Ho	19	ens				
gccctg ttatgg cactga	gtcg ctg cgag ggg aggt gag	ggatgctg gcatctgt gcagccag	attettgtet caacegagte cegteggggg tatatecaga	tggaacggat agctgctggc	gccacggctg ctcttttgag	ctcatccaga	60 120 180 240 249
	<210> 65 <211> 15 <212> DR <213> HG	58	ens				
cactga	ctct ctt	ccatgagt tttcttca	gtcatgaggg ctccagccgg tcttcaaact	ccatccacca	cttaggtgac tattcctggt	aagcacatga tttgaggtaa	60 120 158
	<210> 66 <211> 13 <212> DI <213> Ho	32	ens				
caggto	<400> 60 cccc tto gcagg ago gaagc ca	ccaacccc	tctcaccatc gatcctggat	tcctgatgtg gaggtgagct	cacatcccat ggcggggccg	ggctgtgggc aggtagaggg	60 120 132
	<210> 6' <211> 2: <212> DI	16					

<213> Homo sapiens <400> 67 60 tcttccttcc acctttgtct ccattctacc tgctgtccac tgcagtctga ggacacagac ctgccctacc caccaccca gagggaggcc aacatctaca tggttcctca gaacatcaag 120 ccagcgctcc agcgtaccgc catcgaggtg agccgtccgg gcctgggcgt gggggctggg 180 216 agcagectge cetteceett eetggeecea geettt <210> 68 <211> 263 <212> DNA <213> Homo sapiens <400> 68 cccgggcctt ctgagccact ctcctcattc tgtgtgctta gaatcctggc atggggcctg 60 120 cggaacatga agagttacca gctggccaac atctcctccc ccagcctcgt ggtagagtgt 180 gggggccaga cggtgcagtc ctgtgtcatc aggaacctcc ggaagaaccc caactttgac atotgcacco tottcatgga agtggtgago occacetece tactgtecce ttecagagte 240 263 ctggggctag aagttctaca tgt <210> 69 <211> 249 <212> DNA <213> Homo sapiens <400> 69 60 caggocagtg cgttcttcct cctccaccca gatgctgccc agggaggagc tctactgccc ccccatcacc gtcaaggtca tcgataaccg ccagtttggc cgccggcctg tggtgggcca 120 gtgtaccate cgctccctgg agagettect gtgtgaccec tactcggcgg agagtccate 180 240 cccacagggt ggcccaggta ggggaagggg agatgatggg caggtcaggg aagggggagc 249 ctagggcaa <210> 70 <211> 180 <212> DNA <213> Homo sapiens <400> 70 60 agggggagc cttttgagag agcccctgtc aggcctggat ggctccctcc cctgcagacg atqtqaqcct actcaqtcct ggggaagacg tgctcatcga cattgatgac aaggagcccc 120 tcatecceat ccaggtagga tgggcatect ccagggagge ctgggtcace ttteccetee 180 <210> 71 <211> 211 <212> DNA <213> Homo sapiens <400> 71 tgctgcttgg cgagtcctgt ttctgaaatg gtctctttct ttctacccac tcaggaggaa 60 gagttcatcg attggtggag caaattcttt gcctccatag gggagaggga aaagtgcggc 120 tectacetgg agaaggattt tgacaceetg aaggtaagge etetetteag tetgacagte 180 211 ggtgtgtgtg tgcgtgctgg gcagtgggag a <210> 72 <211> 235 <212> DNA <213> Homo sapiens gttctacttt ctttctgtct cttgtcccct cctctaatcc ccatgtgtgg caggtctatg 60 acacacaget ggagaatgtg gaggeetttg agggeetgte tgaettttgt aacacettea 120 agetgtaceg gggcaagaeg caggaggaga cagaagatee atetgtgatt ggtgaattta 180 aggtaaatcc tcgaagacgt ccctaaccca ggtgggccta agactgtggt gttgg 235 <210> 73 <211> 268

<212> DNA

214

#### 41/68

#### <213> Homo sapiens <400> 73 ggggacacag ccaaaccata tcaacaatga tgataaaata aaattaaccc ttccttcttt tcagggcctc ttcaaaattt atcccctccc agaagaccca gccatcccca tgcccccaag 60 120 acagticcae cagetggceg eccagggace ccaggagtge ttggteegta tetacattgt 180 ccgagcattt ggcctgcagc ccaaggaccc caatggaaag gtaactttct agagccctca 240 268 cctcccaga gtagcaggct caggtaca <210> 74 <211> 200 <212> DNA <213> Homo sapiens <400> 74 tttggaaagt gttttcacag aagtgttttg tctcctcctc cagtgtgatc cttacatcaa 60 gatetecata gggaagaaat cagtgagtga ccaggataac tacateceet gcacgetgga 120 gcccgtattt ggaaagtaaa ttggggcatc ttgggtcttg gggtggagga gccagacagg 180 200 ataacccaca gtctagtggg <210> 75 <211> 263 <212> DNA <213> Homo sapiens <400> 75 cctgttccct tgggtgccct gtgttggctg acattcggga atctgcccct tcctgcagga 60 tgttcgagct gacctgcact ctgcctctgg agaaggacct aaagatcact ctctatgact 120 atgacetect etecaaggae gaaaagateg gtgagaeggt egtegaeetg gagaaeagge tgetgteeaa gtttgggget egetgtggae teceaeagae etaetgtgtg taegtggatg 180 240 263 ggggctggct gcctgcttct ctg <210> 76 <211> 237 <212> DNA <213> Homo sapiens <400> 76 aagcateteg tetatgtett gtgettgete etcagetetg gaeegaacea gtggegggae 60 cageteegee ecteceaget ectecacete ttetgecage ageatagagt caaggeacet 120 gtgtaccgga cagaccgtgt aatgtttcag gataaagaat attccattga agagataggt 180 237 gagetgecae atgaceccaa accatggtgg getetegetg tatecetece tetetea <210> 77 <211> 245 <212> DNA <213> Homo sapiens <400> 77 tetetegett ecceagetee tgeaactttt ttgtgttete tetggggeag aggetggeag 60 gatoccaaac ccacacetgg goccagtgga ggagogtotg gototgcatg tgottcagca 120 gcagggcctg gtcccggagc acgtggagtc acggcccctc tacagccccc tgcagccaga 180 catcgagcag gtaggacctt accettggte ccagagteet egaacteeag aageccaace 240 245 ccagg <210> 78 <211> 214 <212> DNA <213> Homo sapiens <400> 78 60 ggtgcttggt aacagctggt taaatgagaa gggtggggag agaacggacc tgtctccgca ggggaagetg gggaagetge agatgtgggt egacetattt eegaaggeee tggggeggee 120 tggacctccc ttcaacatca ccccacggag agccagaagg tgacttccca gccacaggct 180

ctgagctggg ctgaggggtg gggcgttgca gcct

<210> 79 <211> 229 <212> DNA <213> Homo sapie	ens				
<400> 79 ttcttaaggc cttcccatcc ccaggttttt cctgcgttgt gcctcacggg ggagaagatg tgcctgtcca gcttcccgca	attatctgga agcgacattt	ataccagaga atgtgaaagg	tgtgatcctg gtagggagcc	gatgacctga	60 120 180 229
<210> 80 <211> 261 <212> DNA <213> Homo sapie	ens				
<pre>&lt;400&gt; 80 acgatgtata tactgtgttg tagttggatg attggctttg gggaggtgaa ggcaacttca gcaagtctgt accattgcca cagggcttct aaagttagcc</pre>	aagaacacaa actggaggtt agaaggtcag	gcaaaagaca cattttcccc	gacgtgcatt ttcgactacc	atcgttccct tgccagctga	60 120 180 240 261
<210> 81 <211> 234 <212> DNA <213> Homo sapie	ens				
<400> 81 tgcctctctc taactttgct cttctggagg ctggacaaga ggacaatgac aagttctcct tagaatccca ttctgcacat	ctgagagcaa ttgatgattt	aatcccagca tctggtgatt	cgagtggtgt ttctgggtaa	tccagatctg gcgctattgc	60 120 180 234
<210> 82 <211> 297 <212> DNA <213> Homo sapie	ens				
<400> 82 ggctacaggc tggcagtgat cctgcagctc gatctcaacc ggaccagctg gatgatgctt agtgaagggc tggtggccct tctacttcct ccagccccag	gcatgcccaa tccacccaga gtgtagcaga	gccagccaag atggtttgtg agagggtgag	acagccaaga tccctttttg aagaaaatac	agtgctcctt agcagaaaac tggcggtaag	60 120 180 240 297
<210> 83 <211> 237 <212> DNA <213> Homo sapie	ens				
<pre>&lt;400&gt; 83 cctggttact ctccaggcca cccctcaggg caagctggaa ggcctgctgg ccagggccgg cagtgcccag cccctgagcc</pre>	atgaccttgg gatgagccca	agattgtagc acatgaaccc	agagagtgag taagcttgag	catgaggagc gacccaaggt	60 120 180 237
<210> 84 <211> 252 <212> DNA <213> Homo sapie	ens				
<pre>&lt;400&gt; 84 ccctagtaaa ggatgcccag ttcctgtggt ttacctcccc tgggccatca tcctcttcat tacgccttcc cggtgagcag</pre>	atacaagacc catcctcttc	atgaagttca atcctgctgc	tcctgtggcg tgttcctggc	gcgtttccgg catcttcatc	60 120 180 240

```
252
gggagttcat ca
      <210> 85
      <211> 391
      <212> DNA
      <213> Homo sapiens
      <400> 85
tggctgtgcc tgccccagtg ggatcaccat gggtccctgt ctcctccctc cctccagaac tatgctgcca tgaagctggt gaagcccttc agctgaggac tctcctgccc tgtagaaggg
                                                                          120
geogtggggt eccetecage atgggaetgg cetgeeteet eegeceaget eggegagete
                                                                           180
ctccagacct cctaggcctg attgtcctgc cagggtgggc agacagacag atggaccggc
                                                                          240
ccacattccc agagttgcta acatggaget etgagateac eccacttcca teattteett
                                                                           300
ctccccaac ccaacgcttt tttggatcag ctcagacata tttcagtata aaacagttgg
                                                                           360
                                                                           391
aaccacaaaa aaaaaaaaaa aaaaaaaaaa a
      <210> 86
      <211> 51
      <212> PRT
      <213> Homo sapiens
      <400> 86
Lys Lys Arg Thr Lys Val Ile Lys Asn Ser Val Asn Pro Val Trp Asn
                                       10
                                                            15
Glu Gly Phe Glu Trp Asp Leu Lys Gly Ile Pro Leu Asp Gln Gly Ser
                                   25
Glu Leu His Val Val Lys Asp His Glu Thr Met Gly Arg Asn Arg
          35
Phe Leu Gly
     50
       <210> 87
       <211> 45
       <212> PRT
       <213> Homo sapiens
       <400> 87
Ser Lys Ile Leu Glu Lys Thr Ala Asn Pro Gln Trp Asn Gln Asn Ile
                                        10
  1
Thr Leu Pro Ala Met Phe Pro Ser Met Cys Glu Lys Met Arg Ile Arg
                                   25
              20
Ile Ile Asp Trp Asp Arg Leu Thr His Asn Asp Ile Val
                                40
          35
       <210> 88
       <211> 82
       <212> PRT
       <213> Homo sapiens
       <400> 88
 Gln Ala Arg Asp Leu Ala Ala Met Asp Lys Asp Ser Phe Ser Asp Pro
                                                             15
                                        10
 Tyr Ala Ile Val Ser Phe Leu His Gln Ser Gln Lys Thr Val Val Val
                                                         30
                                    25
               20
 Lys Asn Thr Leu Asn Pro Thr Trp Asp Gln Thr Leu Ile Phe Tyr Glu
                                40
          35
 Ile Glu Ile Phe Gly Glu Pro Ala Thr Val Ala Glu Gln Pro Pro Ser
                                                 60
                           55
 Ile Val Val Glu Leu Tyr Asp His Asp Thr Tyr Gly Ala Asp Glu Phe
                                             75
 Met Gly
       <210> 89
       <211> 79
        <212> PRT
```

<213> Homo sapiens

وَ

#### 44/68

```
<400> 89
Ile Tyr Ile Val Arg Ala Phe Gly Leu Gln Pro Lys Asp Pro Asn Gly
                                      10
Lys Cys Asp Pro Tyr Ile Lys Ile Ser Ile Gly Lys Lys Ser Val Ser
                                  25
                                                      30
             20
Asp Gln Asp Asn Tyr Ile Pro Cys Thr Leu Glu Pro Val Phe Gly Lys
                              40
         35
Met Phe Glu Leu Thr Cys Thr Leu Pro Leu Glu Lys Asp Leu Lys Ile
                         55
Thr Leu Tyr Asp Tyr Asp Leu Leu Ser Lys Asp Glu Lys Ile Gly
                     70
      <210> 90
      <211> 152
      <212> DNA
      <213> Homo sapiens
      <400> 90
acgatgtata tactgtgttg gaaatcttaa tgagaactat tctctaaaaa catgtatgtc
                                                                         60
tagttggatg attggctttg aagaacacaa gcaaaagaca gacgtgcatt atcgttccct
                                                                        120
gggaggtgaa ggcaacttca actggaggtt ca
                                                                        152
      <210> 91
      <211> 56
      <212> DNA
      <213> Homo sapiens
      <400> 91
gtcagtgtcc ttccgattcc ctgtggtgcc agcaccaggg cttctaaagt tagcct
                                                                         56
      <210> 92
      <211> 55
      <212> DNA
      <213> Homo sapiens
tgcctctctc taactttgct tccttgcatc cttctctgtt cctcttccgg gtcag
                                                                         55
      <210> 93
      <211> 68
      <212> DNA
      <213> Homo sapiens
      <400> 93
gtaagcgcta ttgctagaat cccattctgc acatgggggc tgccccagaa cccacactgt
                                                                         60
                                                                         68
gtgtttat
      <210> 94
      <211> 56
      <212> DNA
      <213> Homo sapiens
      <400> 94
ggctacaggc tggcagtgat cgagaaaccc ggccaaaaac cacctctctg ttgcag
                                                                         56
      <210> 95
      <211> 62
       <212> DNA
       <213> Homo sapiens
gtaagtctac ttcctccagc cccagtggag ggcatggggg aagcttcttc catagaaatt
                                                                         60
                                                                         62
qt
       <210> 96
       <211> 68
```

<212> DNA

<	213>	Homo	варіє	ens				
<	400>	96						60
cctggtt	act c	tccag	ggcca	ctgagcagag	ccttcgtgcc	cctaaccaag	tgctctctgt	60 68
ccctca	g							
	210>							
	211>							
	212>		ganie	and				
<	213>	Homo	Bapie	5115				
<	400>	97				accet at acc	cacacteca	59
gtcagtg	CCC E	agccc	ctgag	ccccaatgcc	eacaggeeeg	ggggcacagg	cacageoou	•
<	210>	98						
<	211>	44						
	212>				•			
<	213>	Homo	варт	ens				
<	400>	98						44
ccctagt	aaa q	ggatg	cccag	ttgactccgg	gatetegett	ccag		44
<	210>	99						
	211>							
	212>							
<	213>	Homo	sapi	ens				
<	<400>	99						60
gtgagca	aggc (	ctgac	gacac	tgtggtgggg	gaactctggg	tctaatgggg	gagttcatca	60
<	<210>	100						
	211>							
<	<212>	DNA						
<	<213>	Homo	sapi	ens				
<	<400>	100						
tggctgt	tgcc	tgccc	cagtg	ggatcaccat	gggtccctgt	ctcctccctc	cctccag	57
	<210>	101						
	<211>							
	<212>							
•	<213>	Homo	sapi	.ens				
	<400>	101						
tctctt			ggcca	tag				23
	<210>	102						
	<210 <i>&gt;</i>							
	<212>							
		Homo	sapi	Lens				
	<400>	102						
ctgttc	ctcc	ccato	gtctc	atgg				24
-	<210>	102						
	<210 <i>&gt;</i>							
	<211>							
		Homo	sapi	iens				
	<400>	103						
gctcct			ctctc	g				20
-				_				
	<210>							
	<211>	> 21 > DNA						
		> Homo	sap:	iens				
			_					

<400> 104 gggtcccagc caggagcact g	21
<210> 105 <211> 24 <212> DNA	
<213> Homo sapiens	
<400> 105 cccctctcac catctcctga tgtg	24
<210> 106 <211> 25	
<212> DNA <213> Homo sapiens	
<400> 106	25
tggcttcacc ttccctctac ctcgg	25
<210> 107 <211> 24	
<212> DNA <213> Homo sapiens	
<400> 107	
tcctttggta ggaaatctag gtgg	24
<210> 108 <211> 21	
<212> DNA <213> Homo sapiens	
<400> 108	
ggaagctgga caggcaagag g	21
<210> 109 <211> 27	
<212> DNA	
<213> Homo sapiens	,
<pre>&lt;400&gt; 109 atatactgtg ttggaaatct taatgag</pre>	27
<210> 110	
<211> 21 <212> DNA	
<213> Homo sapiens	
<400> 110 gctggcacca cagggaatcg g	21
<210> 111	
<211> 25 <212> DNA	
<213> Homo sapiens	
<400> 111 ctttgcttcc ttgcatcctt ctctg	25
<210> 112	
<211> 21 <212> DNA	
<213> Homo sapiens	
<400> 112 agccccatg tgcagaatgg g	21

```
<210> 113
        <211> 21
         <212> DNA
        <213> Homo sapiens
        <400> 113
                                                                                           21
 ggcagtgatc gagaaacccg g
        <210> 114
        <211> 21
         <212> DNA
        <213> Homo sapiens
        <400> 114
                                                                                           21
 catgccctcc actggggctg g
        <210> 115
         <211> 21
         <212> DNA
         <213> Homo sapiens
        <400> 115
                                                                                           21
 ggatgcccag ttgactccgg g
         <210> 116
         <211> 21
         <212> DNA
         <213> Homo sapiens
        <400> 116
                                                                                           21
 ccccaccaca gtgtcgtcag g
         <210> 117
         <211> 6240
         <212> DNA
         <213> Homo sapiens
         <400> 117
atgctgaggg tcttcatcct ctatgccgag aacgtccaca cacccgacac cgacatcagc
                                                                                          60
                                                                                        120
gatgectact geteegeggt gtttgeaggg gtgaagaaga gaaccaaagt catcaagaac
agcgtgaacc ctgtatggaa tgagggattt gaatgggacc tcaagggcat ccccctggac
                                                                                        180
cagggetetg agetteatgt ggtggteaaa gaccatgaga egatggggag gaacaggtte
                                                                                        240
ctgggggaag ccaaggtccc actccgagag gtcctcgcca cccctagtct gtccgccagc
                                                                                        300
ttcaatgcc ccctgctgga caccaagaag cagcccacag gggcctcgct ggtcctgcag gtgtcctaca caccgctgcc tggagctgtg ccctgttcc cgccccctac tcctctggag ccctccccga ctctgcctga cctggatgta gtggcagaca caggaggaga ggaagacaca
                                                                                        360
                                                                                        420
                                                                                        480
                                                                                        540
gaggaccagg gactcactgg agatgaggcg gagccattcc tggatcaaag cggaggcccg
ggggctccca ccaccccaag gaaactacct tcacgtcctc cgccccacta ccccgggatc
                                                                                        600
                                                                                        660
aaaaqaaaqc qaaqtqcqcc tacatctaga aagctqctqt cagacaaacc gcaggatttc
cagatcaggg tccaggtgat cgaggggcgc cagctgccgg gggtgaacat caagcctgtg gtcaaggtta ccgctgcagg gcagaccaag cggacgcgga tccacaaggg aaacagccca ctcttcaatg agactcttt cttcaacttg tttgactctc ctggggagct gtttgatgag
                                                                                        720
                                                                                        780
                                                                                        840
                                                                                        900
cccatcttta tcacggtggt agactctcgt tctctcagga cagatgctct cctcggggag
ttccggatgg acgtgggcac catttacaga gagccccggc acgcctatct caggaagtgg
                                                                                        960
                                                                                       1020
ctgctgctct cagaccctga tgacttctct gctggggcca gaggctacct gaaaacaagc
ctttgtgtgc tggggcctgg ggacgaagcg cctctggaga gaaaagaccc ctctgaagac aaggaggaca ttgaaagcaa cctgctccgg cccacaggcg tagccctgcg aggagcccac
                                                                                       1080
                                                                                       1140
                                                                                       1200
ttotgootga aggtottoog ggoogaggao ttgoogoaga tggacgatgo ogtgatggac
aacgtgaaac agatctttgg cttcgagagt aacaagaaga acttggtgga cccctttgtg
                                                                                       1260
gaggtcagct ttgcggggaa aatgctgtgc agcaagatct tggagaagac ggccaaccct cagtggaacc agaacatcac actgcctgcc atgtttccct ccatgtgcga aaaaatgagg
                                                                                       1320
                                                                                       1380
attogtatoa tagactggga cogoctgact cacaatgaca togtggotac cacctacetg
                                                                                       1440
                                                                                       1500
agtatgtega aaatetetge eectggagga gaaatagaag aggageetge aggtgetgte
aagcettega aageeteaga ettggatgae tacetggget teeteeceae tittgggeee
                                                                                       1560
tgctacatca acctctatgg cagtcccaga gagttcacag gcttcccaga cccctacaca
                                                                                       1620
gageteaaca caggeaaggg ggaaggtgtg gettategtg geeggettet geteteectg gagaceaage tggtggagea cagtgaacag aaggtggagg acetteetge ggatgacate
                                                                                       1680
                                                                                       1740
```

ctccgggtgg agaagtacct taggaggcgc aagtactccc tgtttgcggc cttctactca 1800 1860 gccaccatgc tgcaggatgt ggatgatgcc atccagtttg aggtcagcat cgggaactac gggaacaagt tegacatgae etgeetgeeg etggeeteea ecaeteagta cageegtgea 1920 gtctttgacg ggtgccacta ctactaccta ccctggggta acgtgaaacc tgtggtggtg 1980 2040 ctgtcatcct actgggagga catcagccat agaatcgaga ctcagaacca gctgcttggg attgctgacc ggctggaagc tggcctggag caggtccacc tggccctgaa ggcgcagtgc 2100 2160 tecacggagg acgtggacte getggtgget cagetgacgg atgageteat egeaggetge 2220 agccagcete tgggtgacat ccatgagaca ccetetgeca cccacetgga ccagtacetg taccagetge geacecatea cetgagecaa ateactgagg etgecetgge cetgaagete ggecacagtg ageteetge agetetggag caggeggagg actggeteet gegtetgegt 2280 2340 gccctggcag aggagcccca gaacagcctg ccggacatcg tcatctggat gctgcaggga 2400 gacaagcgtg tggcatacca gcgggtgccc gcccaccaag tcctcttctc ccggcggggt 2460 2520 gccaactact gtggcaagaa ttgtgggaag ctacagacaa tctttctgaa atatccgatg gagaaggtgc ctggcgcccg gatgccagtg cagatacggg tcaagctgtg gtttgggctc tctgtggatg agaaggagtt caaccagttt gctgagggga agctgtctgt ctttgctgaa 2580 2640 acctatgaga acgagactaa gttggccctt gttgggaact ggggcacaac gggcctcacc 2700 2760 taccccaagt tttctgacgt cacgggcaag atcaagctac ccaaggacag cttccgcccc 2820 teggeegget ggaeetggge tggagattgg ttegtgtgte eggagaagae tetgeteeat gacatggacg ccggtcacct gagcttcgtg gaagaggtgt ttgagaacca gacccggctt cccggaggcc agtggatcta catgagtgac aactacaccg atgtgaacgg ggagaaggtg 2880 2940 cttcccaagg atgacattga gtgcccactg ggctggaagt gggaagatga ggaatggtcc 3000 acagacetea acegggetgt egatgageaa ggetgggagt atageateae cateceeeg 3060 3120 gagoggaago cgaagcactg ggtocotgot gagaagatgt actacacaca cogacggogg cgctgggtgc gcctgcgcag gagggatctc agccaaatgg aagcactgaa aaggcacagg caggcggagg cggagggcga gggctgggag tacgcctctc tttttggctg gaagttccac ctcgagtacc gcaagacaga tgccttccgc cgccgccgct ggcgccgtcg catggagcca 3180 3240 3300 ctggagaaga cggggcctgc agctgtgttt gcccttgagg gggccctggg cggcgtgatg 3360 3420 gatgacaaga gtgaagatte catgteegte tecacettga getteggtgt gaacagacee acgattteet geatattega etatgggaae egetaceate taegetgeta eatgtaceag geeegggaee tggetgegat ggacaaggae tettttetg atceetatge eategtetee 3480 3540 ttoctgcacc agagecagaa gaeggtggtg gtgaagaaca ceettaacce cacetgggac 3600 cagacgetea tettetaega gategagate titiggegage eggeeacagt tgetgageaa 3660 3720 ccgcccagca ttgtggtgga gctgtacgac catgacactt atggtgcaga cgagtttatg ggtcgctgga ttgtggggg gctgtctggaa cggatgccac ggctggcctg gttcccactg acgaggggca gccagccgtc gggggagctg ctggcctctt ttgagctcat ccagagagag aagccggcca tccaccatat tcctggtttt gaggtgcagg agacatcaag gatcctggat 3780 3840 3900 gagtetgagg acacagacet gecetaceca ceaceceaga gggaggecaa catetacatg 3960 4020 gttcctcaga acatcaagec agegetecag egtacegeca tegagateet ggeatgggge ctgcggaaca tgaagagtta ccagetggee aacateteet eccecageet egtggtagag 4080 tgtgggggcc agacggtgca gtcctgtgtc atcaggaacc tccggaagaa ccccaacttt 4140 gacatetgea ecetetteat ggaagtgatg etgeceaggg aggageteta etgeceecee ateacegtea aggteatega taacegeeag tttggeegee ggeetgtggt gggeeagtgt 4200 4260 4320 accatccgct ccctggagag cttcctgtgt gacccctact cggcggagag tccatcccca cagggtggcc cagacgatgt gagcctactc agtcctgggg aagacgtgct catcgacatt gatgacaagg agccctcat ccccatccag gaggaagagt tcatcgattg gtggagcaaa ttctttgcct ccatagggga gagggaaaag tgcggctcct acctggagaa ggattttgac 4380 4440 4500 accetgaagg tetatgacae acagetggag aatgtggagg cetttgaggg cetgtetgae 4560 4620 ttttgtaaca ccttcaaget gtaccggggc aagacgcagg aggagacaga agatccatct gtgattggtg aatttaaggg cctcttcaaa atttatcccc tcccagaaga cccagccatc 4680 4740 cccatgccc caagacagtt ccaccagctg gccgcccagg gaccccagga gtgcttggtc cgtatctaca ttgtccgagc atttggcctg cagcccaagg accccaatgg aaagtgtgat 4800 cettacatca agatetecat agggaagaaa teagtgagtg accaggataa etacatecee 4860 4920 tgcacgctgg agcccgtatt tggaaagatg ttcgagctga cctgcactct gcctctggag 4980 aaggacctaa agatcactct ctatgactat gacctcctct ccaaggacga aaagatcggt gagacggtcg tcgacctgga gaacaggctg ctgtccaagt ttggggctcg ctgtggactc ccacagacct actgtgtctc tggaccgaac cagtggcggg accagctccg cccctcccag 5040 5100 ctcctccacc tcttctgcca gcagcataga gtcaaggcac ctgtgtaccg gacagaccgt 5160 5220 qtaatqtttc aggataaaga atattccatt gaagagatag aggctggcag gatcccaaac ccacacetgg geccagtgga ggagegtetg getetgeatg tgetteagea geagggeetg gteceggage acgtggagte acggeceete tacageeece tgeagecaga categageag gggaagetge agatgtggt egacetattt ecgaaggee tgggeggee tggaceteee 5280 5340 5400 ttcaacatca ccccacggag agccagaagg tttttcctgc gttgtattat ctggaatacc 5460 5520 agagatgtga teetggatga eetgageete aegggggaga agatgagega eatttatgtg 5580 aaaqqttgga tgattggctt tgaagaacac aagcaaaaga cagacgtgca ttatcgttcc ctgggaggtg aaggcaactt caactggagg ttcattttcc ccttcgacta cctgccagct gagcaagtct gtaccattgc caagaaggat gccttctgga ggctggacaa gactgagagc 5640 5700 aaaatcccag cacgagtggt gttccagatc tgggacaatg acaagttctc ctttgatgat 5760

				49/68			
			1	49/00	و في ا		
aagtget gageage etggege ectgete	tcct aaaa ggca ggcc acct	tggaccagct cagtgaaggg agctggaaat agggccggga ccttcctgtg	cgatctcaac ggatgatgct ctggtggccc gaccttggag tgagcccaac gtttacctcc catcctcttc	tgtgtagcag attgtagcag atgaacccta ccatacaaga atcatcctct	aagagggtga agagtgagca agcttgagga ccatgaagtt tcatcctgct	gaagaaaata tgaggagcgg cccaaggcgc catcctgtgg gctgttcctg	5820 5880 5940 6000 6060 6120 6180 6240
		)> 118					
	<211 <212	l> 13 2> DNA					
	<213	3> Homo sap	Tens				
cgcaa		)> 118 g ctg					13
	<210	> 119					
		L> 12 2> DNA					
		3> Homo sap	iens				
	<400	o> 119					
gagac							12
	<210	0> 120					
	<21	1> 21					
		2> DNA	iona				
	<21.	3> Homo sap	telle				
		0> 120					21
gatct	aacc	c tgctgctca	ie e				
		0> 121					
		1> 21 2> DNA					
		3> Homo sar	piens				
	-40	0> 121					
ctaat		t gcagagcgo	et g				21
		0> 122 1> 21					
	<21	2> DNA					
	<21	3> Homo saj	piens				
		0> 122					21
cctct	cttc	t gctgtctt	ca g				<del></del> –
		.0> 123					
		.1> 21 .2> DNA					
		3> Homo sa	piens				
	-40	00> 123					
tgtgi		t tcaccttc	gt g				21
	<b>~</b> 21	.0> 124					
	<21	1> 21					
		L2> DNA L3> Homo sa	niene				
			Prene				
A		00> 124	aa C				21
tcca		ga aatgeetg	44 C				
		LO> 125					
	<2]	11> 21					

<212> DNA <213> Homo sapi	iens	
<400> 125 aggtatcacc tccaagtgtt	t g	21
<210> 126 <211> 21		
<212> DNA <213> Homo sapi	iens	
<400> 126 taccagette agageteeet	e g	21
<210> 127 <211> 19 <212> DNA		
<213> Homo sapi	Lens	
<400> 127 ttgatcaggg tgctcttgg	:	19
<210> 128 <211> 20		
<212> DNA <213> Homo sapi	iens	
<400> 128 ggagaattgc ttgaacccag	g	20
<210> 129 <211> 22		
<212> DNA <213> Homo sapi	iens	
<400> 129 tggctaatga tgttgaacat	t tt	22
<210> 130 <211> 21		
<212> DNA <213> Homo sap:	iens	
<400> 130 gacccacaag cggcgcctcg	g g	21
<210> 131 <211> 21		
<212> DNA <213> Homo sap	iens	
<400> 131 gaccccggcg agggtggtc	g g	21
<210> 132 <211> 24		
<212> DNA <213> Homo sap	iens	
<400> 132 tgtctctcca ttctccctt	t tgtg	24
<210> 133 <211> 24		
<212> DNA <213> Homo sap	iens	

<400> 133	24
aggacactgc tgagaaggca cctc	24
.040. 404	
<210> 134 <211> 21	
<211> 21 <212> DNA	
<213> Homo sapiens	
<400> 134	21
agtgccctgg tggcacgaag g	21
<210> 135	
<211> 24 <212> DNA	
<213> Homo sapiens	
<400> 135	24
cctacctgca ccttcaagcc atgg	24
<210> 136	
<211> 23	
<212> DNA	
<213> Homo sapiens	
<400> 136	
cagaagagcc agggtgcctt agg	23
<210> 137	
<211> 24	
<212> DNA	
<213> Homo sapiens	
<400> 137	•
ccttggacct taacctggca gagg	24
coccyguodo cadoreggen gargy	
<210> 138	
<211> 21	
<212> DNA	
<213> Homo sapiens	
<400> 138	. 21
cgaggccagc gcaccaacct g	
<210> 139	
<211> 22	
<212> DNA	
<213> Homo sapiens	
<400> 139	22
actgccggcc attcttgctg gg	
<210> 140	
<211> 21	
<212> DNA	
<213> Homo sapiens	
<400> 140	21
ccaggcctca ttagggccct c	21
<210> 141	
<211> 22	
<212> DNA	
<213> Homo sapiens	
<400> 141	22
ctgaagagga gcctggggtc ag	22

õ

	<210> 142	
	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
	<400> 142	
ctgag	atttc tgactcttgg ggtg	24
5-5		
	<210> 143	
	<211> 24	
	<del>-</del>	
	<212> DNA	
	<213> Homo sapiens	
	<400> 143	
aaggt	tetge ceteatgece catg	24
	<210> 144	
	<211> 21	
	<212> DNA	
	<213> Homo sapiens	
	<400> 144	
ctggc	ctgag ggatcagcag g	21
	<210> 145	
	<211> 23	
	<212> DNA	
	<213> Homo sapiens	
	12137 Nome Bapterio	
	<400> 145	
	<400> 145	23
gtgca	tacat acageceaeg gag	23
	<210> 146	
	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
	<400> 146	
		24
gaget	attgg gttggccgtg tggg	2 7
	1010 147	
	<210> 147	
	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
	<400> 147	
accaa	cacgg agaagtgaga actg	24
	<210> 148	
	<211> 26	
	<211> 26 <212> DNA	
	<213> Homo sapiens	
	<400> 148	
ccaca	cttta tttaacgctt tggcgg	26
	<210> 149	
	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
	2210 Homo eattern	
	Z4005 149	
	<400> 149	2
cagaa	ccaaa atgcaaggat acgg	24
	<210> 150	
	<211> 25	
	<212> DNA	

<213> Homo sapiens	
<400> 150	
cttctgattc tgggatcacc aaagg	25
ettergatic egggateres traffs	
<210> 151	
<211> 22	
<212> DNA	
<213> Homo sapiens	
<400> 151	
ggaccgtaag gaagacccag gg	22
ggaccgcaag gaagaaaaay 33	
<210> 152	
<211> 24	
<212> DNA	
<213> Homo sapiens	
<400> 152	0.4
cctgtgctca ggagcgcatg aagg	24
<210> 153	
<211> 22	
<212> DNA	
<213> Homo sapiens	
<400> 153	
gcagacetec cacecaaggg cg	22
<b>30-3</b> -0-0-0-0-0-0-0-0-0-0-0-0-0-0-0-0-0-	
<210> 154	
<211> 24	
<212> DNA	
<213> Homo sapiens	
<400> 154	24
gagacagatg ggggacagtc aggg	24
<210> 155	
<211> 21	
<212> DNA	
<213> Homo sapiens	
<400> 155	21
cctcccgaga gaaccctcct g	
<210> 156	
<211> 21	
<212> DNA	
<213> Homo sapiens	
<400> 156	21
gggagcccag agtccccatg g	
1010> 157	
<210> 157 <211> 21	
C/113 /1	
<212> DNA	
<212> DNA <213> Homo sapiens	
<212> DNA <213> Homo sapiens <400> 157	21
<212> DNA <213> Homo sapiens	21
<212> DNA <213> Homo sapiens <400> 157	21
<212> DNA <213> Homo sapiens <400> 157 gggcctcctt gggtttgctg g	21
<212> DNA <213> Homo sapiens  <400> 157 gggcctcctt gggtttgctg g  <210> 158	21
<212> DNA <213> Homo sapiens <400> 157 gggcctcctt gggtttgctg g	21

<400> 158	21
gcctccccag catcctgccg g	21
<210> 159	
<211> 24	
<212> DNA <213> Homo sapiens	
72137 Nome Baptone	
<400> 159	24
tcactgagcc gaatgaaact gagg	
<210> 160	
<211> 24 <212> DNA	
<213> Homo sapiens	
<pre>&lt;400&gt; 160 tgtggcctga gttcctttcc tgtg</pre>	24
tgtggtttga gttttttt	_
<210> 161	
<211> 24	
<212> DNA <213> Homo sapiens	
<400> 161	24
ggtcaaaggg cagaacgaag aggg	24
<210> 162	
<211> 21	
<212> DNA	
<213> Homo sapiens	
<400> 162	
cccgtccttc tcccagccat g	21
<210> 163	
<211> 21	
<212> DNA	
<213> Homo sapiens	
<400> 163	
ctccctggt tgtccccaag g	21
<210> 164	
<211> 24	
<212> DNA	
<213> Homo sapiens	
<400> 164	
cgacccctct gattgccact tgtg	24
2010\ 16E	
<210> 165 <211> 21	
<211> 21 <212> DNA	
<213> Homo sapiens	
< <b>4</b> 00> <b>1</b> 65	
ggcatcctgc ccttgccagg g	21
<210> 166 <211> 20	
<211> 20 <212> DNA	
<213> Homo sapiens	
4005 166	
<pre>&lt;400&gt; 166 tctgtctccc ctgctccttg</pre>	20

<210> 167	
<211> 21	
<212> DNA	
<213> Homo sapiens	
<400> 167	
cttccctgcc ccgacgccca g	21
effectiget cegacgeous 9	
<210> 168	
<211> 21	
<212> DNA	
<213> Homo sapiens	
1210 Homo Dapania	
<400> 168	
	. 21
cagegeteag gecegtetet e	
<210> 169	
<211> 24	
<212> DNA	
<213> Homo sapiens	
<400> 169	24
tgcataggca tgtgcagctt tggg	24
<210> 170	
<211> 21	
<212> DNA	
<213> Homo sapiens	
<400> 170	
catgcacct ctgccctgtg g	21
Catycacce etgeology y	
-010> 171	
<210> 171	
<211> 21	
<212> DNA	
<213> Homo sapiens	
•	
<400> 171	
	21
agttgagcca ggagaggtgg g	
<210> 172	
<211> 24	
<212> DNA	
<213> Homo sapiens	
CZISZ Momo Bapiciis	
.4005 170	
<400> 172	24
catcaggcgc attccatctg tccg	
<210> 173	
<211> 24	
<212> DNA	
<212> DNA <213> Homo sapiens	
<213> Homo sapiens	
<400> 173	24
agcaggagag cagaagaaga aagg	25
<210> 174	
<211> 22	
<212> DNA	
<213> Homo sapiens	
<400> 174	
gtgtgtcacc atccccaccc cg	22
<210> 175	
<211> 25	
<211> Z3	

<213> Ho	omo sapiens	
<400> 17	75	
caagagatgg gag		25
	, , , ,	
<210> 17		
<211> .23		
<212> DN		
<213> Ho	omo sapiens	
<400> 17	76	
ctgggacatc cgg		23
ccgggacacc cgg	3000030 033	
<210> 17	77	
<211> 22	2	
<212> DN	NA A	
<213> Ho	omo sapiens	
.400. 15		
<400> 17		22
tccaggtagt ggg	gaggeaga gg	
<210> 17	78	
<211> 24		
<212> DN	NA A	
<213> Ho	omo sapiens	
<400> 17		24
teccaetace tgg	gagetgee ttgg	
<210> 17	79	
<211> 2		
<212> Di		
	omo sapiens	
<400> 17		21
ggctctcccc ago	edetedet g	
<210> 18	80	
<211> 24		
<212> DI	NA .	
<213> Ho	omo sapiens	
<400> 18		24
cagagcagca gag	gaeteega eeag	-
<210> 18	81	
<211> 2	1	
<212> DI		
<213> Ho	omo sapiens	
14005 1	0.1	
<400> 10 tagaccccac cto		21
cagaccccac cc.	geceega g	
<210> 1	82	
<211> 2		
<212> D		
<213> H	omo sapiens	
-400s 4s	02	
<400> 10 tecteteatt ge		24
coccoace go		-
<210> 1		
<211> 2		
<212> D	NA .	
<213> H	omo sapiens	

<400> 183	
ttgagagctt gccggggatg g	21
<210> 184	
<211> 24	
<212> DNA	
<213> Homo sapiens	
ZIJV Momo Bupicio	
<400> 184	•
	24
aagtgccaag caatgagtga ccgg	
<210> 185	
<211> 21	
<212> DNA	
<213> Homo sapiens	
<400> 185	
ctcactccca cccaccacct g	21
<210> 186	
<del>-</del>	
<211> 21	
<212> DNA	
<213> Homo sapiens	
<400> 186	
cccaccggcc tctgagtctg c	21
<210> 187	
<211> 24	
<212> DNA	
<213> Homo sapiens	
12137 Homo Baptons	
4400 197	
<400> 187	24
accetaceca agecaggaca agtg	2-3
<210> 188	
<211> 24	
<212> DNA	
<213> Homo sapiens	
-	
<400> 188	
gaatctgcca taaccagctt cgtg	24
<210> 189	
<211> 24	
<212> DNA	
<213> Homo sapiens	
<400> 189	24
tatcacccca tagaggcctc gaag	24
<210> 190	
<211> 24	
<212> DNA	
<213> Homo sapiens	
<400> 190	
cagccactca ctctggcacc tctg	24
cayocacca ceeeggenee coeg	
<210> 191	
<211> 24	
<212> DNA	
<213> Homo sapiens	
<400> 191	
ageceacagt ctctgactct cctg	24

t

ø

<210> 192	•	
<211> 24	·	
<212> DNA		
<213> Homo	anions	
<213/ HOMO	Baptens	
<400> 192		
		24
acatctctca gggtco	cetge tgtg	24
<210> 193		
<211> 21		
<212> DNA		
	anniona	
<213> Homo	adrena	
<400> 193		
	~~~~	21
cctgtgaggg gacgag	ggcag g	21
<210> 194		
	·	
<211> 24		
<212> DNA		
<213> Homo	ganieng	
\213> HOMO	paprens	
<400> 194		
gccctgggta agggat	tacta atto	24
geeergggea aggga	ageog acco	_ 7
<210> 195	•	
<211> 21		
· <del>-</del>		
<212> DNA		
<213> Homo	saniens	
12137 HOMO	odp.c	
<400> 195		
cctgcctggg cctcct	tagat c	21
cordectagg cores	-99	
<210> 196		
<211> 21		
<212> DNA		
<213> Homo	sapiens	
	•	
4400> 106		
<400> 196		
gagggtgatg ggggc	cttag g	21
3 333 3 3 3323		
40105 107		
<210> 197		
<211> 24		
<212> DNA		
<213> Homo	adhrena	
	·	
<400> 197		
		24
gcaatcagtt tgaag	aagga aagg	£ 4
<210> 198		
2011- 04		
<211> 24		
<212> DNA		
<213> Homo	sapiens	
-1207 1101110		
<400> 198		
cacctttgtc tccat	tctac ctgc	24
.0.1.0: 1.0.0		
<210> 199		
<211> 21		
<212> DNA		
<213> Homo	sapiens	
<400> 199		
		21
ctcccagccc ccacg	cccag g	<b>Z</b> 1
	•	
<210> 200		
<211> 24 <212> DNA		

<213> Homo sapiens	
<400> 200	
ctgagccact ctcctcattc tgtg	24
<210> 201 <211> 21	
<211> 21 <212> DNA	
<213> Homo sapiens	
<400> 201	21
tggaagggga cagtagggag g	21
<210> 202	
<211> 22 <212> DNA	
<213> Homo sapiens	
<400> 202	22
ggccagtgcg ttcttcctcc tc	22
<210> 203	
<211> 22 <212> DNA	
<213> Homo sapiens	
<400> 203	
tccctgacct gcccatcatc tc	22
<210> 204	
<211> 21	
<212> DNA <213> Homo sapiens	
<400> 204	
gccctgtca ggcctggatg g	21
<210> 205	
<211> 21	
<212> DNA <213> Homo sapiens	
	·
<400> 205 tgacccaggc ctccctggag g	21
•	
<210> 206 <211> 24	
<212> DNA	
<213> Homo sapiens	
<400> 206	24
ctgaaatggt ctctttcttt ctac	24
<210> 207	
<211> 24 <212> DNA	
<213> Homo sapiens	
<400> 207	
cacaccgact gtcagactga agag	24
<210> 208	
<211> 24	
<212> DNA <213> Homo sapiens	
attac materia	

<400> 208	24
ttgtcccctc ctctaatccc catg	24
<210> 209	
<211> 21	
<212> DNA	
<213> Homo sapiens	
<400> 209	
gggttaggga cgtcttcgag g	21
<210> 210	
<211> 22	
<212> DNA	
<213> Homo sapiens	
<400> 210	
cagccaaacc atatcaacaa tg	22
<210> 211	
<211> 21	
<212> DNA	
<213> Homo sapiens	
<400> 211	
ctggggaggt gagggctcta g	21
<210> 212 <211> 21	
<211> DNA	
<213> Homo sapiens	
<400> 212	
gaagtgtttt gtctcctcct c	21
<210> 213	
<211> 20 <212> DNA	
<213> Homo sapiens	
1220 Monto Department	
<400> 213	20
gcaggcagcc agcccccatc	20
<210> 214	
<211> 21	
<212> DNA	
<213> Homo sapiens	
<400> 214	21
gggtgccctg tgttggctga c	21
<210> 215	
<211> 20	
<212> DNA	
<213> Homo sapiens	
<400> 215	
gcaggcagcc agcccccatc	20
<210> 216 <211> 24	
<211> 24 <212> DNA	
<213> Homo sapiens	
<400> 216	
ctcgtctatg tcttgtgctt gctc	24

<210> 217	
<211> 23	
<212> DNA	
<213> Homo sapiens	
<400> 217	
caccatggtt tggggtcatg tg	g 23
Caccacygue cygycoucy cy	<b>-</b> -
<210> 218	
<211> 21	
<212> DNA	
<213> Homo sapiens	
•	
<400> 218	
tctcgcttcc ccagctcctg c	21
_	
<210> 219	•
<211> 22	
<212> DNA	
<213> Homo sapiens	
1400> 210	
<400> 219	22
tetggagtte gaggaetetg gg	
<210> 220	
<211> 21	
<212> DNA	
<213> Homo sapiens	
12.20	
<400> 220	
agaagggtgg ggagagaacg g	21
<210> 221	
<211> 21	
<212> DNA	
<213> Homo sapiens	
<400> 221	21
cageteagag cetgtggetg g	
<210> 222	
<211> 24	
<211> 24 <212> DNA	
<213> Homo sapiens	<b>,</b>
1210, Homo Dapaone	
<400> 222	
aaggeettee cateetttgg ta	agg 24
33	
<210> 223	
<211> 21	
<212> DNA	
<213> Homo sapiens	
<400> 223	21
acaacccaga gggagcacgg g	
<210> 224	
<210> 224 <211> 25	
<212> DNA	
<213> Homo sapiens	3
<400> 224	
gttgacgatg tatatactgt gt	tgg 25
<210> 225	
<211> 25	
<212> DNA	

<213> Homo sapiens	
<400> 225	
geetetetet aactitgett cettg	25
<210> 226	
<211> 24	
<212> DNA	
<213> Homo sapiens	
<400> 226	
ggctacaggc tggcagtgat cgag	24
<210> 227	
<211> 21	
<212> DNA	
<213> Homo sapiens	
<400> 227	
ttececeatg ecetecaetg g	21
<210> 228	
<211> 24	
<212> DNA	
<213> Homo sapiens	
<400> 228	
agccttcgtg cccctaacca agtg	24
<210> 229	
<211> 21	
<212> DNA	
<213> Homo sapiens	
<400> 229	
ctgtgggcat tggggctcag g	21
<210> 230	
<211> 20	
<212> DNA	
<213> Homo sapiens	
<400> 230	
gccccagtgg gatcaccatg	20
<210> 231	
<211> 21	
<212> DNA	
<213> Homo sapiens	
<400> 231	
atgctggagg ggaccccacg g	21
<210> 232	
<211> 3671	
<212> DNA	
<213> Homo sapiens	
<220>	
<221> CDS	
<222> (418)(3381)	
.400000	
<400> 232	60
tcctggttca agcgattctc tggcctcagc ctcccgagta gctgggatta caggcatgct ccaccaagcc cgggtaattt tgtattttta atagagacgg ggttttgcca tgttggtcag	120
getggteteg aacteetgae etcaggtgat etgeecacet tggeeteca acgtgetgag	180
attacaggca tgagtcactg tgcccggcag agatggtcta attcatatga aagaactctg	240

aaaa ctaa ggag	cctt	at a	gaat	tttt	t tt	tttt	gaga	agc	aaat	tgc	aaat	ttgt	ga τ	agat	ag a	la	300 360 420
ctc Leu	Gly ggg	aga Arg	ttg Leu 5	agc Ser	cta Leu	gag Glu	aaa Lys	gga Gly 10	aga Arg	ttt Phe	gtg Val	aac Asn	cca Pro 15	gga Gly	ggc Gly		468
aga Arg	ggt Gly	aga Arg 20	gat Asp	cca Pro	gga Gly	gag Glu	ggc Gly 25	ggc Gly	gtg Val	atg Met	gat Asp	gac Asp 30	aag Lys	agt Ser	gaa Glu		516
gat Asp	tcc ser 35	atg Met	tcc Ser	gtc Val	tcc Ser	acc Thr 40	ttg Leu	agc Ser	ttc Phe	ggt Gly	gtg Val 45	aac Asn	aga Arg	ccc Pro	acg Thr		564
att Ile 50	tcc Ser	tgc Cys	ata Ile	ttc Phe	gac Asp 55	tat Tyr	ggg ggg	aac Asn	cgc Arg	tac Tyr 60	cat His	cta Leu	cgc Arg	tgc Cys	tac Tyr 65		612
atg Met	tac Tyr	cag Gln	gcc Ala	cgg Arg 70	gac Asp	ctg Leu	gct Ala	gcg Ala	atg Met 75	gac Asp	aag Lys	gac Asp	tct Ser	ttt Phe 80	tct Ser		660
gat Asp	ccc Pro	tat Tyr	gcc Ala 85	atc Ile	gtc Val	tcc Ser	ttc Phe	ctg Leu 90	cac His	cag Gln	agc Ser	cag Gln	aag Lys 95	acg Thr	gtg Val		708
gtg Val	gtg Val	aag Lys 100	aac Asn	acc Thr	ctt Leu	aac Asn	ccc Pro 105	acc Thr	tgg Trp	gac Asp	cag Gln	acg Thr 110	ctc Leu	atc Ile	ttc Phe		756
tac Tyr	gag Glu 115	atc Ile	gag Glu	atc Ile	ttt Phe	ggc Gly 120	gag Glu	ccg Pro	gcc Ala	aca Thr	gtt Val 125	gct Ala	gag Glu	caa Gln	ccg Pro		804
ccc Pro 130	agc Ser	att Ile	gtg Val	gtg Val	gag Glu 135	ctg Leu	tac Tyr	gac Asp	cat His	gac Asp 140	act Thr	tat Tyr	ggt Gly	gca Ala	gac Asp 145		852
gag Glu	ttt Phe	atg Met	ggt Gly	cgc Arg 150	tgc Cys	atc Ile	tgt Cys	caa Gln	ccg Pro 155	agt Ser	ctg Leu	gaa Glu	cgg Arg	atg Met 160	PIO		900
cgg Arg	ctg Leu	gcc Ala	tgg Trp 165	ttc Phe	cca Pro	ctg Leu	acg Thr	agg Arg 170	ggc Gly	agc Ser	cag Gln	ccg Pro	tcg Ser 175	Gly	gag Glu		948
ctg Leu	ctg Leu	gcc Ala 180	tct Ser	ttt Phe	gag Glu	ctc Leu	atc Ile 185	cag Gln	aga Arg	gag Glu	aag Lys	ccg Pro 190	Ala	atc Ile	cac His		996
cat His	att Ile 195	cct Pro	ggt Gly	ttt Phe	gag Glu	gtg Val 200	cag Gln	gag Glu	aca Thr	tca Ser	agg Arg 205	atc Ile	ctg Leu	gat Asp	gag Glu		1044
tct Ser 210	Glu	gac	aca Thr	gac Asp	ctg Leu 215	Pro	tac Tyr	cca Pro	cca Pro	Pro 220	GIn	agg Arg	gag Glu	gcc Ala	aac Asn 225	l .	1092
atc Ile	tac Tyr	atg Met	gtt Val	cct Pro 230	Gln	aac Asn	atc Ile	aag Lys	cca Pro 235	Ala	ctc Leu	cag Gln	cgt Arg	acc Thr 240	. Ala	: L	1140

atc Ile	gag Glu	atc Ile	ctg Leu 245	gca Ala	tgg Trp	ggc Gly	ctg Leu	cgg Arg 250	aac Asn	atg Met	aag Lys	agt Ser	tac Tyr 255	cag Gln	ctg Leu	1188
gcc Ala	aac Asn	atc Ile 260	tcc Ser	tcc Ser	ccc Pro	agc Ser	ctc Leu 265	gtg Val	gta Val	gag Glu	tgt Cys	999 Gly 270	ggc Gly	cag Gln	acg Thr	1236
gtg Val	cag Gln 275	tcc Ser	tgt Cys	gtc Val	atc Ile	agg Arg 280	aac Asn	ctc Leu	cgg Arg	aag Lys	aac Asn 285	ccc Pro	aac Asn	ttt Phe	gac Asp	1284
atc Ile 290	tgc Cys	acc Thr	ctc Leu	ttc Phe	atg Met 295	gaa Glu	gtg Val	atg Met	ctg Leu	ccc Pro 300	agg Arg	gag Glu	gag Glu	ctc Leu	tac Tyr 305	1332
tgc Cys	ccc Pro	ccc Pro	atc Ile	acc Thr 310	gtc Val	aag Lys	gtc Val	atc Ile	gat Asp 315	aac Asn	cgc Arg	cag Gln	ttt Phe	ggc Gly 320	cgc Arg	1380
cgg Arg	cct Pro	gtg Val	gtg Val 325	ggc Gly	cag Gln	tgt Cya	acc Thr	atc Ile 330	cgc Arg	tcc Ser	ctg Leu	gag Glu	agc Ser 335	ttc Phe	ctg Leu	1428
tgt Cys	gac Asp	ccc Pro 340	tac Tyr	tcg Ser	gcg Ala	gag Glu	agt Ser 345	cca Pro	tcc Ser	cca Pro	cag Gln	ggt Gly 350	ggc Gly	cca Pro	gac Asp	1476
gat Asp	gtg Val 355	agc Ser	cta Leu	ctc Leu	agt Ser	cct Pro 360	ggg Gly	gaa Glu	gac Asp	gtg Val	ctc Leu 365	atc Ile	gac Asp	att Ile	gat Asp	1524
gac Asp 370	aag Lys	gag Glu	ccc Pro	ctc Leu	atc Ile 375	ccc Pro	atc Ile	cag Gln	gag Glu	gaa Glu 380	gag Glu	ttc Phe	atc Ile	gat Asp	tgg Trp 385	1572
tgg Trp	agc Ser	aaa Lys	ttc Phe	ttt Phe 390	gcc Ala	tcc Ser	ata Ile	gly ggg	gag Glu 395	agg Arg	gaa Glu	aag Lys	tgc Cys	ggc Gly 400	tcc Ser	1620
tac Tyr	ctg Leu	gag Glu	aag Lys 405	gat Asp	ttt Phe	gac Asp	acc Thr	ctg Leu 410	aag Lys	gtc Val	tat Tyr	gac Asp	aca Thr 415	cag Gln	ctg Leu	1668
gag Glu	aat Asn	gtg Val 420	gag Glu	gcc Ala	ttt Phe	gag Glu	ggc Gly 425	ctg Leu	tct Ser	gac Asp	ttt Phe	tgt Cys 430	aac Asn	acc Thr	ttc Phe	1716
aag Lys	ctg Leu 435	tac Tyr	cgg Arg	ggc Gly	aag Lys	acg Thr 440	cag Gln	gag Glu	gag Glu	aca Thr	gaa Glu 445	gat Asp	cca Pro	tct Ser	gtg Val	1764
att Ile 450	Gly	gaa Glu	ttt Phe	aag Lys	ggc Gly 455	ctc Leu	ttc Phe	aaa Lys	att Ile	tat Tyr 460	ccc Pro	ctc Leu	cca Pro	gaa Glu	gac Asp 465	1812
cca Pro	gcc Ala	atc Ile	ccc Pro	atg Met 470	Pro	cca Pro	aga Arg	cag Gln	ttc Phe 475	cac His	cag Gln	ctg Leu	gcc Ala	gcc Ala 480	cag Gln	1860
gga Gly	ccc Pro	cag Gln	gag Glu 485	Cys	ttg Leu	gtc Val	egt Arg	atc Ile 490	Tyr	att Ile	gtc Val	cga Arg	gca Ala 495	ttt Phe	ggc Gly	1908
ctg Leu	cag Gln	Pro 500	aag Lys	gac Asp	ccc Pro	aat Asn	gga Gly 505	Lys	tgt Cys	gat Asp	cct Pro	tac Tyr 510	Ile	aag Lys	atc Ile	1956

tcc Ser	ata Ile 515	ggg Gly	aag Lys	aaa Lys	tca Ser	gtg Val 520	agt Ser	gac Asp	cag Gln	gat Asp	aac Asn 525	tac Tyr	atc Ile	ccc Pro	tgc Cys	2004
acg Thr 530	ctg Leu	gag Glu	ccc Pro	gta Val	ttt Phe 535	gga Gly	aag Lys	atg Met	ttc Phe	gag Glu 540	ctg Leu	acc Thr	tgc Cys	act Thr	ctg Leu 545	2052
cct Pro	ctg Leu	gag Glu	aag Lys	gac Asp 550	cta Leu	aag Lys	atc Ile	act Thr	ctc Leu 555	tat Tyr	gac Asp	tat Tyr	gac Asp	ctc Leu 560	ctc Leu	2100
tcc Ser	aag Lys	gac Asp	gaa Glu 565	aag Lys	atc Ile	ggt Gly	gag Glu	acg Thr 570	gtc Val	gtc Val	gac Asp	ctg Leu	gag Glu 575	aac Asn	agg Arg	2148
ctg Leu	ctg Leu	tcc Ser 580	aag Lys	ttt Phe	gly ggg	gct Ala	cgc Arg 585	tgt Cys	gga Gly	ctc Leu	cca Pro	cag Gln 590	acc Thr	tac Tyr	tgt Cys	2196
gtc Val	tct Ser 595	gga Gly	ccg Pro	aac Asn	cag Gln	tgg Trp 600	cgg Arg	gac Asp	cag Gln	ctc Leu	cgc Arg 605	ccc Pro	tcc Ser	cag Gln	ctc Leu	2244
ctc Leu 610	cac His	ctc Leu	ttc Phe	tgc Cys	cag Gln 615	cag Gln	cat His	aga Arg	gtc Val	aag Lys 620	gca Ala	cct Pro	gtg Val	tac Tyr	cgg Arg 625	2292
aca Thr	gac <b>A</b> sp	cgt Arg	gta Val	atg Met 630	ttt Phe	cag Gln	gat Asp	aaa Lys	gaa Glu 635	tat Tyr	tcc Ser	att Ile	gaa Glu	gag Glu 640	ata Ile	2340
gag Glu	gct Ala	ggc Gly	agg Arg 645	atc Ile	cca Pro	aac Asn	cca Pro	cac His 650	ctg Leu	ggc Gly	cca Pro	gtg Val	gag Glu 655	gag Glu	cgt Arg	2388
ctg Leu	gct Ala	ctg Leu 660	cat His	gtg Val	ctt Leu	cag Gln	cag Gln 665	cag Gln	ggc Gly	ctg Leu	gtc Val	ccg Pro 670	gag Glu	cac His	gtg Val	2436
gag Glu	tca Ser 675	cgg Arg	ccc Pro	ctc Leu	tac Tyr	agc Ser 680	ccc Pro	ctg Leu	cag Gln	cca Pro	gac Asp 685	atc Ile	gag Glu	cag Gln	Gly	2484
aag Lys 690	ctg Leu	cag Gln	atg Met	Trp	gtc Val 695	Asp	Leu	Phe	Pro	Lys	Ala	ctg Leu	ggg Gly	cgg Arg	cct Pro 705	2532
gga Gly	cct Pro	ccc Pro	ttc Phe	aac Asn 710	Ile	acc Thr	cca Pro	cgg Arg	aga Arg 715	gcc Ala	aga Arg	agg Arg	ttt Phe	ttc Phe 720	ctg Leu	2580
cgt Arg	tgt Cys	att Ile	atc Ile 725	Trp	aat Asn	acc Thr	aga Arg	gat Asp 730	Val	atc Ile	ctg Leu	gat Asp	gac Asp 735	Leu	agc Ser	2628
ctc Leu	acg Thr	ggg Gly 740	Glu	aag Lys	atg Met	agc Ser	gac Asp 745	Ile	tat Tyr	gtg Val	aaa Lys	ggt Gly 750	Trp	atg Met	att Ile	2676
ggc Gly	ttt Phe 755	Glu	gaa Glu	cac His	aag Lys	caa Gln 760	Lys	aca Thr	gac	gtg Val	cat His 765	Tyr	cgt Arg	tcc Ser	ctg Leu	2724

Ŷ

gga Gly 770	ggt Gly	gaa Glu	ggc Gly	aac Asn	ttc Phe 775	aac Asn	tgg Trp	agg Arg	ttc Phe	att Ile 780	ttc Phe	ccc Pro	ttc Phe	gac Asp	tac Tyr 785	2772
ctg Leu	cca Pro	gct Ala	gag Glu	caa Gln 790	gtc Val	tgt Cys	acc Thr	att Ile	gcc Ala 795	aag Lys	aag Lys	gat Asp	gcc Ala	ttc Phe 800	tgg Trp	2820
agg Arg	ctg Leu	gac Asp	aag Lys 805	act Thr	gag Glu	agc Ser	aaa Lys	atc Ile 810	cca Pro	gca Ala	cga Arg	gtg Val	gtg Val 815	ttc Phe	cag Gln	2868
atc Ile	tgg Trp	gac Asp 820	aat Asn	gac Asp	aag Lys	ttc Phe	tcc Ser 825	ttt Phe	gat Asp	gat Asp	ttt Phe	ctg Leu 830	ggc Gly	tcc Ser	ctg Leu	2916
cag Gln	ctc Leu 835	gat Asp	ctc Leu	aac Asn	cgc Arg	atg Met 840	ccc Pro	aag Lys	cca Pro	gcc Ala	aag Lys 845	aca Thr	gcc Ala	aag Lys	aag Lys	2964
tgc Cys 850	tcc Ser	ttg Leu	gac Asp	cag Gln	ctg Leu 855	gat Asp	gat Asp	gct Ala	ttc Phe	cac His 860	cca Pro	gaa Glu	tgg Trp	ttt Phe	gtg Val 865	3012
tcc Ser	ctt Leu	ttt Phe	gag Glu	cag Gln 870	aaa Lys	aca Thr	gtg Val	aag Lys	ggc Gly 875	tgg Trp	tgg Trp	ccc Pro	tgt Cys	gta Val 880	gca Ala	3060
gaa Glu	gag Glu	ggt Gly	gag Glu 885	aag Lys	aaa Lys	ata Ile	ctg Leu	gcg Ala 890	ggc Gly	aag Lys	ctg Leu	gaa Glu	atg Met 895	acc Thr	ttg Leu	3108
gag Glu	att Ile	gta Val 900	gca Ala	gag Glu	agt Ser	gag Glu	cat His 905	gag Glu	gag Glu	cgg Arg	cct Pro	gct Ala 910	ggc Gly	cag Gln	ggc Gly	3156
cgg Arg	gat Asp 915	gag Glu	ccc Pro	aac Asn	atg Met	aac Asn 920	cct Pro	aag Lys	ctt Leu	gag Glu	gac Asp 925	cca Pro	agg Arg	cgc Arg	ccc Pro	3204
gac Asp 930	Thr	tcc Ser	ttc Phe	ctg Leu	tgg Trp 935	ttt Phe	acc Thr	tcc Ser	cca Pro	tac Tyr 940	Lys	acc Thr	atg Met	aag Lys	ttc Phe 945	3252
atc Ile	ctg Leu	tgg Trp	Arg	Arg	ttc Phe	Arg	Trp	Ala	Ile	Ile	Leu	ttc Phe	Ile	atc Ile 960	Leu	3300
ttc Phe	atc Ile	ctg Leu	ctg Leu 965	Leu	ttc Phe	ctg Leu	gcc Ala	atc Ile 970	Phe	atc Ile	tac Tyr	gcc Ala	ttc Phe 975	ccg Pro	aac Asn	3348
tat Tyr	gct Ala	gcc Ala 980	Met	aag Lys	ctg Leu	gtg Val	aag Lys 985	Pro	ttc Phe	agc Ser	tga	ggac	tct	cctg	ccctgt	3401
gac ttt	gctc	ctc	caga cact cccc	ccto ccca aacc	ct a	ggcc ttgc cgct	tgat taac tttt	t gt a tg t gg	cctg gagc	ccag	ggt aga	gggc tcac	aga ccc	caga actt	gctcgg cagatg ccatca ataaaa	3461 3521 3581 3641 3671

<210> 233 <211> 988 <212> PRT <213> Homo sapiens

<400> 233 Met Leu Gly Arg Leu Ser Leu Glu Lys Gly Arg Phe Val Asn Pro Gly Gly Arg Gly Arg Asp Pro Gly Glu Gly Gly Val Met Asp Asp Lys Ser Glu Asp Ser Met Ser Val Ser Thr Leu Ser Phe Gly Val Asn Arg Pro Thr Ile Ser Cys Ile Phe Asp Tyr Gly Asn Arg Tyr His Leu Arg Cys Tyr Met Tyr Gln Ala Arg Asp Leu Ala Ala Met Asp Lys Asp Ser Phe Ser Asp Pro Tyr Ala Ile Val Ser Phe Leu His Gln Ser Gln Lys Thr Val Val Val Lys Asn Thr Leu Asn Pro Thr Trp Asp Gln Thr Leu Ile Phe Tyr Glu Ile Glu Ile Phe Gly Glu Pro Ala Thr Val Ala Glu Gln Pro Pro Ser Ile Val Val Glu Leu Tyr Asp His Asp Thr Tyr Gly Ala Asp Glu Phe Met Gly Arg Cys Ile Cys Gln Pro Ser Leu Glu Arg Met Pro Arg Leu Ala Trp Phe Pro Leu Thr Arg Gly Ser Gln Pro Ser Gly Glu Leu Leu Ala Ser Phe Glu Leu Ile Gln Arg Glu Lys Pro Ala Ile His His Ile Pro Gly Phe Glu Val Gln Glu Thr Ser Arg Ile Leu Asp Glu Ser Glu Asp Thr Asp Leu Pro Tyr Pro Pro Pro Gln Arg Glu Ala Asn Ile Tyr Met Val Pro Gln Asn Ile Lys Pro Ala Leu Gln Arg Thr Ala Ile Glu Ile Leu Ala Trp Gly Leu Arg Asn Met Lys Ser Tyr Gln Leu Ala Asn Ile Ser Ser Pro Ser Leu Val Val Glu Cys Gly Gln Thr Val Gln Ser Cys Val Ile Arg Asn Leu Arg Lys Asn Pro Asn Phe Asp Ile Cys Thr Leu Phe Met Glu Val Met Leu Pro Arg Glu Glu Leu Tyr Cys Pro Pro Ile Thr Val Lys Val Ile Asp Asn Arg Gln Phe Gly Arg Arg Pro Val Val Gly Gln Cys Thr Ile Arg Ser Leu Glu Ser Phe Leu Cys Asp Pro Tyr Ser Ala Glu Ser Pro Ser Pro Gln Gly Gly Pro Asp Asp Val Ser Leu Leu Ser Pro Gly Glu Asp Val Leu Ile Asp Ile Asp Asp Lys Glu Pro Leu Ile Pro Ile Gln Glu Glu Phe Ile Asp Trp Trp Ser Lys Phe Phe Ala Ser Ile Gly Glu Arg Glu Lys Cys Gly Ser Tyr Leu Glu Lys Asp Phe Asp Thr Leu Lys Val Tyr Asp Thr Gln Leu Glu Asn Val Glu Ala Phe Glu Gly Leu Ser Asp Phe Cys Asn Thr Phe Lys Leu Tyr Arg Gly Lys Thr Gln Glu Glu Thr Glu Asp Pro Ser Val Ile Gly Glu Phe Lys Gly Leu Phe Lys Ile Tyr Pro Leu Pro Glu Asp Pro Ala Ile Pro Met Pro Pro Arg Gln Phe His Gln Leu Ala Ala Gln Gly Pro Gln Glu Cys Leu Val Arg Ile Tyr Ile Val Arg Ala Phe Gly Leu Gln Pro Lys Asp Pro Asn Gly Lys Cys Asp Pro Tyr Ile Lys Ile Ser Ile Gly Lys Lys Ser Val Ser Asp Gln Asp Asn Tyr Ile Pro 

J

_	530		Glu			535					540				
Leu 545	Pro	Leu	Glu	Lys	Asp 550	Leu	Lys	Ile	Thr	Leu 555	Tyr	Asp	Tyr	Asp	Leu 560
Leu	Ser	Lys	Asp	Glu 565		Ile	Gly	Glu	Thr 570	Val	Val	Asp	Leu	Glu 575	Asn
Arg	Leu	Leu	Ser 580		Phe	Gly	Ala	Arg 585	Сув	Gly	Leu	Pro	Gln 590	Thr	Tyr
Сув	Val	Ser 595	Gly	Pro	Asn	Gln	Trp 600		Asp	Gln	Leu	Arg 605	Pro	Ser	Gln
Leu	Leu 610	His	Leu	Phe	Cys	Gln 615		His	Arg	Val	Lys 620	Ala	Pro	Val	Tyr
Arg 625	Thr	Asp	Arg	Val	Met 630		Gln	Asp	Lys	Glu 635	Tyr	Ser	Ile	Glu	Glu 640
Ile	Glu	Ala	Gly	Arg 645		Pro	Asn	Pro	His 650	Leu	Gly	Pro	Val	Glu 655	Glu
Arg	Leu	Ala	Leu 660		Val	Leu	Gln	Gln 665	Gln	Gly	Leu	Val	Pro 670	Glu	His
Val	Glu	Ser 675	Arg	Pro	Leu	Tyr	Ser 680	Pro	Leu	Gln	Pro	Asp 685	Ile	Glu	Gln
Gly	Lys 690	Leu	Gln	Met	Trp	Val 695	Asp	Leu	Phe	Pro	Lys 700	Ala	Leu	Gly	Arg
705	_		Pro		710					715					720
Leu	_	_	Ile	725					730					735	
			Gly 740					745					750		
	_	755	Glu				760					765			
	770		Glu			775					780				
785			Ala		790					795					800
			Asp	805					810					815	
			Asp 820					825					830		
		835	yab				840					845			
_	850		Leu			855					860				
865			Phe		870					875					880
			Gly	885					890					895	
			Val 900					905					910		
_	_	915					920					925			
	930		Ser			935					940				
945			Trp		950					955					960
			Leu	965					970				Ala	Phe 975	Pro
Asn	Tyr	Ala	Ala 980		ГÀв	Leu	Val	Lys 985		Phe	Ser				

## INTERNATIONAL SEARCH REPORT

International application No. PCT/US99/19395

	SSIFICATION OF SUBJECT MATTER	:					
\	C12N 15/11, 15/00; C07K 16/00 536/23.1, 435/440, 530/387.1						
	International Patent Classification (IPC) or to both n	ational classification and IPC					
B. FIEL	DS SEARCHED						
Minimum do	ocumentation searched (classification system followed	by classification symbols)					
U.S. :	536/23.1, 435/440, 530/387.1						
Documentati	ion searched other than minimum documentation to the	extent that such documents are included	in the fields searched				
Electronic d	ata base consulted during the international search (nar	ne of data base and, where practicable,	search terms used)				
BIOSIS, C	CAPLUS, EMBASE, ESBIOBASE, LIFESCI, MEDLN rms: dysferlin, lgmd2b						
C. DOC	UMENTS CONSIDERED TO BE RELEVANT						
Category*	Citation of document, with indication, where app	ropriate, of the relevant passages	Relevant to claim No.				
x	WEILER et al. Limb-girdle muscul Myopathy in an aboriginal Canadian ki		32,35				
	segregate with the same haplotype. A	merican Journal of Human	•				
	Genetics. October 1996, Vol.59, pages 872-878, especially page 873.						
x	KOENIG et al. Complete cloning		32-33,36				
	ary genomic organization of d individuals. Cell. 31 July lly pages 511-513.						
X Furth	ner documents are listed in the continuation of Box C.	See patent family annex.					
• Sp	secial categories of cited documents:	*T* later document published after the indicate and not in conflict with the app	emational filing date or priority				
	cument defining the general state of the art which is not considered be of particular relevance	the principle or theory underlying th	e invention				
	rlier document published on or after the international filing date	"X" document of particular relevance; the considered novel or cannot be considered when the document is taken alone	ne claimed invention cannot be seed to involve an inventive step				
ci	ted to establish the publication date of another citation or other ecial reason (as specified)	"Y" document of particular relevance; the considered to involve an inventive	ne claimed invention cannot be				
m	ocument referring to an oral disclosure, use, exhibition or other cans	combined with one or more other su- being obvious to a person skilled in	ch documents, such combination				
th	cument published prior to the international filing date but later than e priority date claimed	*&* document member of the same pater					
Date of the	actual completion of the international search	Date of mailing of the international se					
17 NOVI	EMBER 1999	13 JAN 20	000				
Name and Commission Box PCT	mailing address of the ISA/US oner of Patents and Trademarks	Authorized officer	De for				
Washingto	on, D.C. 20231	Stephen Siu	qu				
I Facsimile	No. (703) 305-4242	Telephone No. (703) 308-0196	•				

Form PCT/ISA/210 (second sheet)(July 1992)\*

International application No. PCT/US99/19395

		Γ
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No
X,P  Y,P	Database GenCore version 4.5, Compugen Ltd., No. AI128455, 'NCI-CGAP, National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index', Unpublished, 27 October 1998	1,6,12  7,14,16
X  Y	Database GenCore version 4.5, Compugen Ltd., No. R41062, WAYE, M.M.Y. et al. 'Gene expression of adult human heart as revealed by random sequencing of cDNA library,' Miami Winter Biotechnol. Symp. Proc. 6,90, 16 May 16, 1995.	1, 6, 11-12  7, 14
X  Y	Database GenCore version 4.5, Compugen Ltd., No. AA718275, Marra et al, 'The WashU-HHMI Mouse EST Project', Unpublished, 29 December 1997.	1, 6, 11-12 7, 14
Y	BASHIR et al. Genetic and physical mapping at the limb-girdle muscular dystrophy locus (LGMD2B) on chromosome 2p. Genomics. April 1996, Vol.33, pages 46-52, especially page 47.	32,36
Y	MOREIRA et al. The seventh form of autosomal recessive limb- girdle muscular dystrophy is mapped to 17q11-12. American Journal of Human Genetics. July 1997, Vol. 61, pages 151-159, entire document.	32, 35
Y	Database GenCore version 4.5, Compugen Ltd., No. R76778, HILLIER et al., 'The WashU-Merck EST Project', Unpublished, 06 June 1995.	7, 14
A,E	AHLBERG et al. Genetic Linkage of Welander Distal Myopathy to chromosome 2p13. Annals of Neurology. September 1999, Vol. 46, No.3, pages 399-404, especially page 400.	37, 39
A,E	BITTNER et al. Dysferlin deletion in SJL mice (SJL-Dysf) defines a natural model for limb girdle muscular dystrophy 2B. Nature Genetics. October 1999, Vol. 23, pages 141-142, especially page 141.	40
A,P	BASHIR et al. A gene related to Caenorhabditis elegans spermatogenesis factor fer-1 is mutated in limb-girdle muscular dystrophy type 2B. Nature Genetics. September 1998, Vol 20, pages 37-42.	1-53
A,E	Matsuda et al. Dysferlin is a surface membrane-associated protein that is absent in Miyoshi Myopathy. Neurology 22 September 1999, Vol. 53, No. 5, pages 1119-1122, especially pages 1119-1120.	40

Form PCT/ISA/210 (continuation of second sheet)(July 1992)★

ntemational	application	No.
PCT/US99/	19395	

# INTERNATIONAL SEARCH REPORT

	Ation). DOCUMENTS CONSIDERED TO BE RELEVANT	Relevant to claim No
Category*	Citation of document, with indication, where appropriate, of the relevant passages	
A,P	LIU et al. Dysferlin, a novel skeletal muscle gene, is mutated in Miyoshi Myopathy and limb girdle muscular dystrophy. Nature Genetics. September 1998, Vol. 20, pages 31-36.	1-54

Form PCT/ISA/210 (continuation of second sheet)(July 1992)\*